DOCUMENT RESUME

ED 021 308 24 EA 001 237

By- Cook, Desmond L.

THE IMPACT OF THE HAWTHORNE EFFECT IN EXPERIMENTAL DESIGNS IN EDUCATIONAL RESEARCH. FINAL REPORT.

Ohio State Univ., Columbus.

Spons Agency- Office of Education (DHEW), Washington, D.C. Bureau of Research.

Report No- P- 1757

Bureau No- BR- 5- 0726

Pub Date Jun 67

Contract- OEC- 3- 10- 041

Note- 213p.

EDRS Price MF- $1.00 HC- $8.60

Descriptors- *BEHAVIORAL SCIENCE RESEARCH, EDUCATIONAL EXPERIMENTS, *EDUCATIONAL RESEARCH, FIELD STUDIES, LITERATURE REVIEWS, REACTIVE BEHAVIOR, *RESEARCH DESIGN, *RESEARCH METHODOLOGY, TABLES (DATA), *VALIDITY

Identifiers- *Hawthorne Effect

Project objectives included (1) establishing a body of knowledge concerning the role of the Hawthorne effect in experimental designs in educational research, (2) assessing the influence of the Hawthorne effect on educational experiments conducted under varying conditions of control, (3) identifying the major components comprising the effect, and (4) suggesting methodological approaches to control the possible influence on the effect in educational experiments. Methodology and procedures involved (1) establishing a field research situation in which selected variables were introduced and manipulated as possible cues to generate the effect, and (2) analyzing completed published studies to determine if results of such studies could be explained on the basis of the possible presence of the Hawthorne effect. Field research findings showed no systematic relation to the achievement gains to cuing variables or to awareness of experimentation. Analysis of the literature revealed no systematic relationship between study results and control procedures employed. (Author/TT)

5- 0726
24

FINAL REPORT
Project No. 1757
Contract No. OE-3-10-041

THE IMPACT OF THE HAWTHORNE EFFECT IN EXPERIMENTAL DESIGNS IN EDUCATIONAL RESEARCH

June 1967

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EA 001 237
ED021308

THE IMPACT OF THE HAWTHORNE EFFECT IN EXPERIMENTAL DESIGNS IN EDUCATIONAL RESEARCH

Project No. 1757
Contract No. OE-3-10-041

Desmond L. Cook

June 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

The Ohio State University
Columbus, Ohio

Contents

Acknowledgments iv

Chapter I - Introduction 1

The Problem. 1
Relevant Literature. 3
Objectives of the Study. 12
A Conceptualization of the Hawthorne effect Phenomenon 14
Organization of the Report 17

Chapter II - Methodology and Procedures. 18

Field Research Phase. 18
Experimental Design. 19
Selection of Experimental Sample 23
Orientation of Participating Teachers. 26
Instrumentation and Data Collection. 29
Literature Analysis Methodology 37
Research Areas Investigated. 37
Abstracting Procedure. 39

Chapter III - Field Research Results 43

Introduction 43
Cues, Awareness, and Achievement 43
Duration of Study. 56
Teacher Interview Results. 60

Chapter IV - Literature Analysis Results 63

Control Procedures for the Hawthorne Effect 63
Definitions and Suggested Controls 63
Control Procedures as Related to Experimental Results. 70
Study Duration and Results. 80
Mechanical Changes and Study Results. 94

Chapter V - Discussion, Conclusions, and Implications. . . 99

Introduction 99
Discussion 99
Conclusions. 112
Implications 120

Chapter VI - Summary 125

The Problem. 125
Procedure. 126
Results. 127
Conclusions. 129
Implications 129

References 131

Bibliography 139

Appendix A - Instructions to Teachers. A-1

Appendix B - Development of the Awareness Inventory. . . . B-1

Appendix C - Literature Analysis Materials C-1

Part 1 - Major Areas of Research. C-1
Part 2 - Abstract Summary Form. C-6
Part 3 - Abstracting Procedure - Experimental Teaching Methods Studies C-7

This report has been prepared in compliance with the publication _Instructions for Preparing Substantive Reports Based on Research Projects Supported by the Office of Education, Bureau of Research_; issued August 18, 1966.

Acknowledgments

A research investigation of the scope and magnitude as that reported in this document could not possibly have been undertaken and concluded without the assistance of many persons and agencies. The contributions of such persons and agencies should be noted so that their efforts will be recognized.

Appreciation is expressed first to the Cooperative Research Program of the U. S. Office of Education for providing the funds to conduct the project. Without such support, the activities reported herein would have taken much longer to accomplish.

The contributions of Dr. David L. Clark, formerly Associate Dean, College of Education, The Ohio State University, to the development of the original proposal and his continuing suggestions during the course of the project were of immense value.

Special acknowledgment is made of the suggestions and comments received from Dr. Julian Stanley, University of Wisconsin, and Dr. Leonard Feldt, University of Iowa, with regard to the experimental design and statistical analyses of the data respectively. Recognition of their contributions, however, should not be construed to mean that the design and statistical analyses employed had their full approval.

Dr. Lewis Coon served in the position of Associate for Mathematics and was responsible for conducting the field research phase of the total project. In addition to carrying out an in-service program for a two year period, he was responsible for most of the tasks associated with the analysis of data and the preparation of a report dealing with this aspect of the project.

Several persons contributed to the literature analysis phase of the total project. Collectively, their contributions facilitated this total activity. Among the persons working on this phase of the project were Walter Calinger, William Gephart, Judy Gerken, Jo Ann Adrion King, Fred Lawrence, and Penny Tricket. Fred Lawrence developed most of the abstracting procedures used in the literature analysis. Jo Ann Adrion King pulled together most of the materials relating to the literature analysis into a comprehensive report which served as the basis for the chapters presenting this phase of the project.

Initial development of the inventory to measure awareness of experimental participation herein was undertaken by Fred Lawrence and Jo Ann Adrion King. The development of the final instrument was the contribution of Herbert Muktarian.

A special debt of gratitude is owed to the Columbus Public School System, Dr. Harold Eibling, Superintendent, for granting the project staff permission to use selected classrooms for a two year period of time with a minimum of administrative difficulties. Concurrent with this acknowledgment is a special note of thanks for the contributions of the participating teachers, many of whom were required to learn the new mathematics to fulfill their role in the classroom.

Such a project as this would not have moved far without the talents and skills of the project secretaries. Mrs. Kay Taylor helped to initiate the project and was followed by Miss Barbara Kinney who served most of the total project time. Mrs. Marilyn Mamrick typed most of the final draft material as well as the final manuscript. Their contributions are deeply appreciated.

Appreciation is also expressed to Dr. Egon Guba, formerly Director, Bureau of Educational Research and Service, The Ohio State University, both for providing sufficient released time to conduct the project and for his many valuable comments and suggestions.

Despite all of this help, the principal investigator must, and does, accept full responsibility for the final product.

ERRATA

Page	Line	Correction
5	10	*reaction's* should be *reactions*
17	1	*results* should be *result*
17	23	insert after the word *definition* "are referred to Kaplan's statement that strict definition..."
32	9	*fro* should be *for*
42	12	second *from* should be *as*
48	footnote 3	insert *certain* between with and other
49	Table 7	Heading "n" should read "final n"
50	4th line from bottom	change *the* to *no*
94	8	change *prime objective* to *general question*
96	27	delete first sentence "Each ... in a study"
115	10	*approved* should be *appeared*
121	5th line from bottom	should read *would not be*
123	8	delete *in*
130	13	*logical biases* should be *logic*
157	--	*Kranfer* should be *Kanfer*
B-1	footnote 1	*structive* should be *interview*
B-13	2nd line from bottom	*Varimox* should be *Varimax*
B-21	8	should read "mean summated and scale scores"

Chapter 1

Introduction

The Problem

The term "Hawthorne effect" has been made popular by researchers in the social and behavioral sciences ever since the time of the now classic illumination experiments which were conducted at the Hawthorne plant of the Western Electric Company in the middle 1920's (66). As generally used, this concept refers to unanticipated but beneficial effects produced in experimental situations. Such effects are said to be caused by the subject in the experiment realizing that he was a participant in an experiment and thus the object of special attention, whether real or imagined. Such awareness generally is said to have a positive facilitating effect on the subject's performance during the duration of the experimental situation.

Despite the rather common use of the Hawthorne effect concept by educational researchers, the phenomenon appears not to have had any direct systematic investigation and, with a relatively few exceptions, has not penetrated the literature of educational research. In view of the continued and increased interest in and expenditures for educational research, a lack of knowledge about this concept in educational experimentation may be causing substantial loss of time, effort, and money through invalidation of research results on the one hand or overly elaborate, ineffective, or inefficient control mechanisms on the other. Thus, the present state of knowledge about the nature and role of the Hawthorne effect in educational experimentation does not permit researchers to arrive at intelligent decisions with regard to its importance and/or control.

Even if an assumption is made by a researcher on the basis of the information presently available that the Hawthorne effect is worth serious consideration in the development of his research design, the social-psychological components of the effect are not sufficiently well enough understood to allow for the election of efficient and effective control mechanisms. If indirect cues (e.g., pretesting) produce the effect, the researcher is in a far different situation than if direct cues (informing subjects), increase subject performance.

The lack of information about the Hawthorne effect places a large number of educational research studies, particularly methods studies conducted in classroom settings, into the following uncomfortable categories:

1. Significant differences in favor of experimental groups in situations where the Hawthorne effect was not controlled may be disregarded or dismissed on this basis alone.

2. No significant difference studies, where both the control and experimental groups were cued to the experimental nature of their situation or provided with unusual opportunities for obtaining attention, may be invalidated on the assumption that the effect of the treatment was obliterated by the more powerful effect of the experimental setting.

3. Studies that did attempt to control for the reactive effect may be questioned on the basis that the wrong component of the effect was controlled or that the effect is not important enough to justify elaborate control.

The situation into which many educational studies would be placed because of the reactive effect occurring between experimental arrangements and the subject's subsequent participation was stated well by Clark (11) in his discussion of the costs associated with bringing about qualitative gains in education.

> It is probably fair to state that essentially all educational experiments of the past seventy years have not been able to disentangle the effect of enthusiasm from the effects of the experimental procedure and content. If during the next seventy years this is accomplished, it may then be possible to assess with confidence the factors that will improve the quality of education. Until such procedures are adopted, however, it is going to be extraordinarily difficult to determine the impact of specific content or method on quality. (11, p. 49).

If the above statement has validity, it would seem imperative that some systematic study be made regarding the Hawthorne effect and its role in educational research.

The primary goal of project described in this report was to attempt to build up a body of information about the Hawthorne effect which could provide a basis for the subsequent development of a theory adequate to explain the effect. It was further hoped that the investigations made would provide information which would allow educational researchers to arrive at more informed decisions with regard to handling or controlling the effect in their experimental designs. The specific objectives and questions to be answered during the course of the project appear in a subsequent section of this chapter.

In order to provide a setting for the actual research activities undertaken during the course of the project, some attention should be given to the role that the Hawthorne effect is said to play in educational research, the original development of the concept, and previous investigations which were influential in the formulation of the general topic under investigation as the project moved along.

Relevant Literature

A systematic review of discussions about the Hawthorne effect is rather difficult to present because the phenomenon is not consistently referred to in educational and other literature by its familiar name. A complete review of relevant available literature at this point is premature because a substantial portion of the total project time was devoted to an examination of published research studies including direct reference to the use of the concept by persons engaged in research.

The research and other relevant literature on the Hawthorne effect concept which was examined did tend to fall into several general categories. Within each of the categories, a large number of position papers and research reports were found to exist. A detailed summary of each category would constitute a substantially large portion of this report. Consequently, an attempt has been made only to highlight the areas by referring to selected publications which give orientation and focus to the categories. Bibliographies relating to each of the categories have been prepared and are included as the Bibliography to this report. Readers interested in any one or all of the categories are directed to the Bibliography for further information regarding the category of concern.

The remaining part of this section is devoted to a brief overview of (1) studies relating to the original Hawthorne research, (2) the social-psychology of experimentation, (3) awareness by subjects of experimental participation, and (4) literature in the field of education relating to the topic.

Original Hawthorne Studies. The basic works relating to the Hawthorne effect and its relationship to experimental treatments are those of Snow (76) and Roethlisberger and Dickson (66). Generally, persons interested in the Hawthorne effect read the latter reference. The article by Snow, however, describes the original series of illumination studies of which the Hawthorne studies were a continuation. Snow's concluding remarks in his paper provided the starting point for the subsequent Hawthorne plant research.

> Any investigation attempting to evaluate definitely the effect of illumination or some such influence must take the greatest of pains to control or eliminate all factors but the one being studied. Many of them can be controlled or eliminated, but the one great stumbling block remaining is the problem of the psychology of the human individual. (76, p. 274).

The Roethlisberger and Dickson report of the research is probably the best known of the several reports relating directly to the original experiments conducted at the Hawthorne plant. The behavior of workers in experimental illumination situations, that of continued high output under varying lighting conditions, was not predicted. By means of an intensive personal interview program, the investigators were able to discover that the workers considered their situation to be one of special treatment or attention from management. The investigators attributed the unexpected production results to the existence of this latter phenomenon only after carefully examining possible alternative hypotheses which included wage incentives, rest pauses, and other factors. Interestingly, the Hawthorne personnel engaged in the project chose to focus further efforts upon the securing of more information about worker's attitudes and feelings with regard to their work situation. A subsequent development was a personal relations program which was the forerunner of the "human relation" programs now existing in many industrial and non-industrial situations. The problem of experimental methodology contamination apparently did not become a major focus of interest to them.

Numerous reports and interpretations of the Hawthorne findings have been prepared over the succeeding years. Among the more outstanding are those by Mayo (51), Pennock (59), Homans (34), Landesberger (42), and Blum (4).

Social Psychology of Experimentation. While the original Hawthorne researchers focused their attention on the human relations program, researchers in the social science areas began to develop interest in the nature of a subject's reaction to experimental activities and procedures and the relation of these reaction's to the subsequent behavior or performance in the experimental situation. It should not be assumed that the report of the Hawthorne studies prompted this concern directly. Rather its appearance during the emergence of the areas of industrial and social psychology no doubt influenced some of the work on research methodology being developed in these areas.

Numerous examples of this concern over subject-situation interaction are evident in many fields. Reports by Crespi (18) and Star and Hughes (77) in the area of public opinion polling demonstrate the positive effect pre-interviewing can have upon follow-up responses. Lana (41) on the other hand, reported that pretesting for attitudes did not sensitize the individual with regard to subsequent attitudes. Studies by Hovland and others (36) revealed that in some situations the subject-situation interaction may be inhibiting in that the subject becomes nonresponsive or responds in an unanticipated manner to the treatment condition. Herzog (32) discussed, in a monograph on guides for evaluation of research in psychotherapy, additional instances of treatment contamination due to possible influence of sensitization on the part of subjects.

Of more relevance than the above research to the general concern of this project is the research relating directly to the social psychology of the experimental situation. This topic has been deemed of sufficient importance that symposia have been held on this topic at annual meetings of the American Psychological Association and the American Educational Research Association.

The respective works of Rosenthal on experimenter-bias and Orne in the area of subject's demands to have their needs met in the experimental situation are worthy of note.

Rosenthal's efforts have concentrated on the influence that the experimenter's wishes and desires have on the outcome of the experimental situation. He has demonstrated in a variety of situations that the experimenter can have his hypothesis validated because of his biases and their interaction with participating subjects, even with animals. His work included the finding of similar results when surrogates (e.g., graduate assistants) were used. A summary of the major findings of several years of research by Rosenthal appears in The American Scientist for June 1963 (69).

The work of Orne has dealt largely with the phenomenon of subject's needs being met in the experimental situation. Orne presents evidence that subjects tend to demand a purpose or reason from the experimental situation, even if a situation is created which has no purpose. Orne indicates that various kinds of cues exist which direct the subject to operate in a certain manner in a specific situation. These various cues are referred to as demand characteristics. Probably the best account of Orne's work appears in the November 1962 issue of The American Psychologist (55).

The combined works of Rosenthal and Orne strongly influenced the thinking of the project staff during the course of the project. While each research effort focused on a different aspect of the experimental situation, the experimenter in one case and the subject in the other, the findings are highly relevant to the general problem of studying the roles of the student and the researcher in a classroom experimental situation.

While not directly involved in actual research, Reicken (63) has stated very well the need to investigate the subject experiment interaction problem if research is to be conducted successfully in the social and behavioral sciences. His general proposition is that the typical experimental situation is a somewhat one-sided operation. The experimenter alone knows what is to be done and what information is to be presented to the subject at what time. The subject, on the other hand, initially possesses none of this information but as he proceeds through the experimental situation he tries to secure more and more meaning. As he does so, his cooperation and responses are modified. Thus, Reicken proposes that a type of subject--experimenter negotiation takes place during the course of an experiment. The subject is also faced with what Reicken refers to as a deutero-problem. On the one hand, he is faced with ascertaining the task to be done and thus meeting the experimenter's approval yet at the same time possessing certain personal needs which call for him to do well (e.g., to protect himself) in the experimental situation.

Awareness of Experimental Participation. If one is to accept the general interpretation of the Hawthorne effect that a subject's performance is favorably influenced by knowledge of participation in an experimental situation through being the object of special attention, then it would seem imperative that research relating to an individual's awareness of things being done to him should be examined. During the course of the project, the concept of subject awareness became a predominant factor in deriving an operational definition of the Hawthorne effect for use in the project and thus directing much of the research effort.

Two general areas of research provide evidence regarding the subject's awareness of experimental situations. One area is that of learning studies and the other is that of psychotherapeutic situations.

Adams (1) has reviewed much of the research relating to learning with awareness and points out that evidence on this phenomenon is generally inconclusive and unverifiable except in the case of studies of stimulus discrimination. Krasner (39) reviewed 31 studies in the area of verbal conditioning from several points of view, including "awareness" and setting. He noted that in only about half of the studies were the subjects able to detect the reinforcing contingency, thus supporting the findings of Adams. Matarazzo and others (50) attempted to isolate the awareness effect as a variable and relate it to performance in verbal conditional experiments in learning. A highly relevant aspect of Matarazzo's work to the present investigation was the post-experimental classification of subjects by levels of awareness as to their knowledge of the purposes and procedures of the experiment. Awareness as defined for their research was observed to be positively related to the subject being favorably conditioned.

In addition to the works of Krasner and Matarazzo cited above, the reader is referred particularly to two additional references in this area. The first is a report of a symposium relating to the research and interpretation of awareness and behavior edited by Eriksen (26). The second is a summary of the research on awareness and its implication for the counseling situation in education prepared by Adrion (2).

One research study directly related to this project was undertaken by Lawrence (44). He attempted to develop instrumentation which would provide a quantitative assessment of awareness of participation in an experimental situation in education. Since the study was conducted for only a very short period of time (two weeks) and as an adjunct to a regular classroom situation, the findings were not conclusive although there was some evidence obtained that students could become aware of their participation in an experimental situation, albeit working under the handicaps noted.

The general conclusion to be drawn from the above research done largely with adults in various situations is that in some cases subjects could indicate their awareness of special experimental conditions but at the same time there were difficulties encountered in developing adequate means to assess this awareness. No evidence was encountered that children of an elementary school age could ascertain their participation in an experimental situation. Such evidence would be necessary if one is going to suggest that many of the studies conducted at the elementary school level, where many outstanding results are obtained, could be attributed to the student's awareness of being the object of special attention.

The Hawthorne effect in Education. Although there is recognition in the literature of educational research as to the existence and possible confounding involvement of the Hawthorne effect in experimental designs, the most frequent reference is usually in terms of a post-experimental apology relative to the possible role of the phenomenon in explaining the success of an experimental group or the appearance of no difference existing between "methods." This point was well made by Smith (75) in his review of why many projects were not approved under the Cooperative Research Program of the U. S. Office of Education. Further, the employment of the concept does not meet with any degree of consistency having been variously referred to as the novelty effect, the halo effect, attention, motivation, and similar synonyms (15).

The principal investigator of this project summarized the early work on the origin of the concept and some of its interpretations plus suggestions for control as they have appeared in the educational research literature in the December 1962 issue of the Phi Delta Kappan (12).[1] Although referred to in a Phi Delta Kappa symposium on educational research design (67), the concept was not defined. Campbell and Stanley (10) in the Handbook of Research on Teaching chose to include the phenomenon under a more general term of "interaction" when discussing factors affecting the external validity of experimental designs. An examination of standard introductory texts on research methodology for education for direct reference to the phenomenon, its impact on experimental designs or suggestions for its control, reveals only occasional reference to the concept, often without elaboration.

---

[1] A detailed review of the content presented in this article has been purposefully omitted in this review since the material is available in published form.

The infrequent reference to the concept in educational research literature directly does not mean that educators are not unaware of its existence. If anything, they perhaps may be only too well aware of its role in the conduct and results of research. Some examples of this concern are presented below.

In the excellent study on the relation of cost to quality of education noted above, Clark (11) refers generally to the concept in terms of teacher enthusiasm for the new method resulting in a transference over to the students causing them to accomplish more than the traditional or control groups. After summarizing selected research studies in several areas of education, Clark is prompted to indicate the significance of this teacher enthusiasm factor.

> The wide range of experimental results in the schools would seem to confirm much the same conclusion. Experiments that have emphasized phonics have brought improvements in reading; experiments that have emphasized the meaning of words but with less emphasis on phonics have also brought increases in reading ability. Comparable experiments have been conducted in all major subject matter fields. Again, the most reasonable interpretation is that there is strong experimental interest-enthusiasm factor at work (11, p. 48).

A somewhat similar interpretation is given by Cronbach (19) in his discussion of problems encountered in curriculum evaluation in order to bring about needed course improvements.

> In an educational experiment, it is difficult to keep pupils unaware they are an experimental group. And it is quite impossible to neutralize the biases of the teacher as those of the doctor are neutralized in the double-blind design. It is thus never certain whether any observed advantage is attributable to the educational innovation as such or to the greater

> energy that teachers and students put forth when a method is fresh and experimental. (19, p. 237).

In his recent book on Innovation in Education (53), Miles refers to the fact that not all educators look negatively upon the operation of the Hawthorne effect in the educational system. Miles goes on to suggest that those who favor capitalizing upon the phenomenon have, in fact, missed the boat.

> And, in one educational meeting after another, deliberate use of the "Hawthorne effect" is advocated--"So long as you do something new, the kids will learn more"--in blissful ignorance of the fact that the work of Roethlisberger and Dickson showed that sheer "newness" was not the issue. Rather alterations in decision-making structure, the development of groups norms, styles of formal and informal leadership, and the perceived reward structure were all factors at work in enhancing (and restricting!) output at the Hawthorne plant. (53, p. 11).

In a footnote to the above comment, Miles states a belief that work being done on the concept by the present project was not correctly oriented since an emphasis was being given only to the popular meaning of the Hawthorne effect. Miles goes on to state that the common definition of the phenomenon completely disregards the aspects of the social system which were, in fact, responsible for the changes noted. In a subsequent portion of his book, Miles returns to the relationship of the Hawthorne effect and its operation in a social system. In his interpretation, an educational experiment is a temporary system, and as such, may have heightened meaning for those involved in it.

> Part of the Hawthorne effect seems to consist of this heightened sense of significance; since the system is temporary its members regard the situation as special, as marked off in some way from the ordinary run of activities in the permanent systems....Heightened significance may explain the fact that almost any educational experiment will show increments in learning over that appearing in associated control groups; such gain often recedes as the experiment becomes less temporary and more durable part of an on-going permanent system. (53, pp. 461-2).

This emphasis on the role enthusiasm plays is given additional emphasis by Scriven in a paper dealing with the methodology of evaluation (72), Scriven places the concept under a more generic concept labeled "disturbance effects" as may be noted in the following statement:

> The enthusiasm difficulty here is simply an example of what we might call disturbance effects (italics mine), of which the placebo effect in medicine and the Hawthorne effect in industrial and social psychology are well-known instances. In each case we are interested in finding out the effects of a certain factor, but we cannot introduce the factor into the experimental situation without producing a disturbance which itself may be responsible for the observed changes. In the drug field the disturbance consists in the act of giving the patient something which he considers to be a drug, something which does not ordinarily happen to him, and consequently may produce effects of its own, quite apart from the effects of the drug. In the Hawthorne effect, the disturbance is the disruption of e.g. conditions which may suggest to the worker that he is the subject of special study and interest, and this may lead to improved output, not the physical changes in the environment that are the intended parameters under study. (72, p. 42).

The above quotations serve to point up the general importance of the phenomenon investigated in the present study, whatever its fundamental nature might be, to both research methodologists and persons actually conducting research. While recognized as an important variable in educational research activities, there appears to have been little direct investigation of the phenomenon which would be of assistance to persons desiring to control for the phenomenon in research.

Not all persons share the concern of the research methodologist and the researcher with regard to the contaminating influence of the Hawthorne effect in research. As noted by Miles, there are those persons who take the general position that if the phenomenon does, in fact and irregardless of its nature, produce better learning or achievement on the part of the student, then we ought to incorporate the Hawthorne effect into all educational methods and/or curriculum study in a deliberate manner. Change, novelty, attention, and similar variables would then be considered to be a part of every modification in education. In one sense, it would allow educators to take the position that the current emphasis on educational change is good if for no other reason than the sake of change. An educator could find it difficult to argue against such a position if there is true and valid interest in improving the educational situation for students to maximize the learning. In short, the ends justify the means.

The principal investigator of this project has taken the position that while we want the best education possible for students, it is equally necessary that we have an understanding of the means by which we achieve the ends. Since the Hawthorne effect possibly operates to contaminate both the means and ends, it is necessary to study the phenomenon to see exactly, or at least with more precision than now exists, what role the phenomenon plays in the research and development activities education undertakes to produce better means designed to achieve the desired ends.

Objectives of the Study.

The overall objective of the study was an attempt to build up a body of information and knowledge about the Hawthorne effect in educational research which, in turn, might provide a basis for the later development of a theory adequate to explain the phenomenon. It should be understood that this objective did not include the development of a theory about the phenomenon as a direct product of the investigations contained in this report. This position was taken on the basis of a general principle derived from the history of science that the first task in establishing a new area of scientific inquiry is an organization of what is already known. It was hoped that the activities of the project might provide further information which would permit educational researchers to make better decisions with regard to the handling of the phenomenon in experimental designs.

General Objectives.

Within the overall objective stated above, three general objectives guided the work of the project.

A. To assess the relative influence of the Hawthorne effect on educational experiments conducted under varying conditions of control.

B. To identify the major components which comprise the Hawthorne effect.

C. To suggest effective and efficient methodological approaches to control or account for any influence which the Hawthorne effect might have on educational experiments.

General Questions. A series of general questions developed from the above three general objectives guided the work of the project. Each of these questions is presented below followed by several specific questions which appeared to be derived from the more general question.

A. Could the Hawthorne effect be identified as the variable accounting for all or a major portion of the difference, or lack of difference, noted between experimental and control groups in selected educational research studies on educational methods? Specific questions under this question were as follows:

1. In completed educational research studies of similar content and approach, were there differences in relative control and experimental group achievement results and study conclusions when varying approaches had been used by the investigators to control for the effect?

2. When taken in combination, were there groups of completed educational research studies which have produced similar results for different methodological approaches when the common element of the studies has been the construction of the design? For example, do reading methods studies show more consistency of results on the basis of whether the adaptation (i.e., phonics vs. word sight) is experimental or control than on the basis of what the experimental method is employed?

3. Could an experimental situation be developed or contrived within which the results or findings of the methods studies varied as a consequence of the way in the Hawthorne effect was controlled or not controlled in the experiment?

4. Can different results be identified, either through the analysis of completed studies or the construction of an experimental situation, on the basis of (a) the duration of the experiment or (b) the mechanistic nature of the experimental and control group settings.

B. Could the component elements of the so-called Hawthorne effect be identified? Within this general question, the following more specific questions were developed?

1. Was the subject's awareness of participation in an experiment sufficient to create differences in achievement results in control and experimental groups?

2. Were the relationships, if any, derived from the preceding question of the same direction and magnitude where direct cues are used as contrasted with situations using indirect cues?

C. From the analysis of information secured to answer questions from A and B above, could the relative efficiency and effectiveness of control mechanisms be inferred?

For the above objective, specific questions were developed as follows:

1. If no overt or purposeful cues are provided to the control group, can the investigator assume the Hawthorne effect is not operating within the control group?

2. To what extent, if any, are placebo treatments necessary as control mechanisms?

3. Can extended duration of the experiment be used as a self-control for the Hawthorne effect?

A Conceptualization of the Hawthorne effect Phenomenon.

The preceding sections of this chapter have outlined the nature of the problem, reviewed some relevant literature, and set forth the general objectives and specific questions for the investigation. It would seem imperative, in view of the relatively unexplored nature of the phenomenon under study, that some definition or conceptualization of the phenomenon under study be presented at this time so the reader can be properly oriented to the balance of the report. The purpose of this section is to present a conceptualization which guided the work of the investigation.[2] Such an approach is somewhat at variance with the more traditional presentation of definitions of terms used in a study report since such presentations require a careful delimitation of concepts. It was deemed unwise to provide a strong delimiting definition in view of the basic exploratory nature of the study. While the conceptualization presented was formed early in the study, it reflects the incorporation of ideas from the literature review noted above. Its presentation follows the ideas set forth by Kaplan in the

---

[2] The presentation here is a briefer version of a paper titled "The Hawthorne Effect: Fact or Artifact" presented at The 1966 Annual Meeting of The American Educational Research Association.

Conduct of Inquiry (37) because of the logical manner by which he suggests one arrives at the meaning of a concept for scientific inquiry.

As Kaplan has noted, any concept is composed of a group of conceptions with a conception being essentially the meaning of the concept in a particular use. The processes of classification and categorization are employed in order to establish which particular conception is being talked about. It would be possible, for example, to discuss the Hawthorne effect in terms of the particular types of experimental situations within which the research on instructional methodology takes place. In this case, one would pay attention to the experimental design being used and its possible accompanying physical changes. One could also discuss the Hawthorne effect in terms of the results or findings of such methodological studies. Here, the concern would be primarily with the achievement test results. A third possible conception would focus around an interpretation of any results as set forth by the original investigator. Such conceptions of the Hawthorne effect did not seem to be useful approaches in establishing a working conception of the Hawthorne effect for research purposes because they do not deal with the role of the subject in the study but only with its external features.

The conception which appeared most useful for the study of Hawthorne effect related to the processes which were introduced and went on during the course of the methodological investigation with the primary concern focusing upon meaning of such processes (e.g. the experimental situation) for the participating subject. Such a conception of the Hawthorne effect focusing upon the effects on subjects of processes of experimentation receives support from the research relating to the social psychology of experimentation noted earlier in this chapter. The conception of the Hawthorne effect in terms of the effects of process enabled us, as Kaplan has so well stated, to mark out the path by which we might move most freely in logical space as well as to mark the category which would tell us more about our subject matter than any other category or classification. It was therefore posited that the Hawthorne effect was a substantive term, not a notational term, whose conception focuses on the process of investigation as represented by the social psychology of experimentation.

A scientific term is given meaning by referring to a set of experiences that lead to the conception. The experiences refer largely to direct or indirect observations, or observables as discussed by Kaplan. What are the observables (i.e., experiences) that form the basis of the conception of Hawthorne effect? In one sense, it might be said that the Hawthorne effect is directly observable in the sense that one could look at achievement test results. It is probably more appropriate to think of it in terms of being an indirect observable rather than a direct observable because we cannot truly see the Hawthorne effect. As an indirect observable, it enables us to establish a connection between what has been observed (i.e., results) and what the term signifies. To some, this would suggest a hypothetical construct or intervening variable but the utilization of these terms creates problems in view of some controversy existing about the use of these terms in the field of scientific inquiry (49).

While it might be debated for some time as to whether or not the Hawthorne effect is a construct or theoretical term, the status assigned to a term depends on the function or use being made of it in a specific context of inquiry. The principal point to be emphasized is that however a theoretical term is to be analyzed and brought into connection with an empirical base, it is indispensable to the actual pursuit of scientific inquiry. Regardless of how the term is described, there is still need to provide further specific meaning to the conception presented above.

In developing a meaning for the concept Hawthorne effect, it should be remembered that a certain openness of meaning will always exist. It should also be remembered that a demand for premature closure by giving a precise definition might well interfere with further inquiry into the concept. A failure to structure a situation as well as one would like does not imply that no inquiry made in that situation is really scientific.

The specification of meaning for a theoretical term can be achieved by the procedures of description and example. A description of the Hawthorne effect would focus upon the unexplained measurable behavior on the part of a subject in an experiment not accountable to any controlled variable.

In a published paper on the phenomenon, the principal investigator proposed a definition of the Hawthorne effect as follows:

> The Hawthorne effect is a phenomenon characterized by a cognitive awareness on the part of the subjects of special treatment created by artificial experimental conditions. It becomes confounded with the independent

> variable under study with the subsequent results of either facilitating or inhibiting the dependent variables under study and leading to spurious conclusions. (12, p. 118).

The essential element of this definition, or conceptualization, of the Hawthorne effect is awareness on the part of the subject. Since awareness is essentially a noun, there is a need to provide more precision in this definition by answering the question Aware of what? To answer this question, it must be realized that almost all instructional methodology experiments involve or exhibit certain experimental conditions which are changes from what has been in existence. To the extent that these changes are not a part of the natural environment of the subject, they can be referred to as artificial. It is also perhaps prudent at this time to point out that no attempt has been made to indicate which set of artificial experimental conditions would create more or less awareness than another set, nor was there any attempt to indicate which element of any one set would contribute a greater weight than another.

It is readily admitted that the conceptualization presented above is perhaps not precise enough at this stage, nor reliable enough for many persons. Those persons who feel the need for a more precise or strict definition is not a precondition of scientific inquiry but actually its culmination. The investigation reported herein was conducted using the Hawthorne effect concept as presented above.

Organization of the Report.

The purpose of this chapter was to outline the background of the study and present the reasons why it was undertaken. The degree to which the project staff was successful in accomplishing the general and specific objectives of the study is contained in the balance of this report.

Chapter II describes the methodological procedures in the two main phases of project activity--the literature analysis and the field experimentation. Chapter III presents the results of the data from the field experimentation, while Chapter IV presents the results of the literature analysis. Chapter V presents a discussion of the results along with conclusions and implications from the total effort. Chapter VI presents a summary of the total report. Appropriate appendices have been included to supply details which would burden the reader if incorporated in the text. The reader is referred to such appendices at appropriate places in the text.

Chapter II

Methodology and Procedures

An analysis of the objectives as set forth in Chapter I indicated that the methodology and procedures employed in the study would be divided into two different but interrelated sets of activities. Further, accomplishment of objectives required that the information obtained from the two major activities be brought back together at the end of the study. The two primary procedures involved a field research phase and a library or literature analysis phase. The purpose of this chapter is to describe the activities carried out under each of these two phases. The methodology and procedures for the field research phase is presented first followed by those for the literature analysis.

The procedural details for the total project were planned and controlled on a time schedule by use of the Program Evaluation and Review Technique, or PERT (14). The considerations involved in establishing the original network are described in a report by Coon (16). The use of PERT in planning and controlling the work flow of the project for a period of approximately three years was found to be a very valuable tool for the management of the project by the principal investigator as well as serving as an excellent communication tool for the project staff.

I. Field Research Phase

The field research phase was undertaken to secure information which might be useful in answering the general objectives of the investigation and most specifically with regard to Questions 3 and 4 under general question A, Questions 1 and 2 of general question B, and Question 3 under general question C.

The general purpose of the field undertaking was to determine the extent to which the Hawthorne effect, as defined, could be generated by introducing selected variables into a typical classroom instructional situation. The variables selected for introduction were to be derived after careful consideration of the Hawthorne effect concept, previous studies on methods and curriculum research in classroom settings, and practical considerations of what would be possible in a cooperating school situation.

Several considerations were primary to the project staff during the planning for the field study. One consideration was the desire not to create a situation which in and of itself would generate the very effect that was under study.

The second was that any experimental situation which would be developed should be as similar as possible in its purpose, scope, and implementation to extant studies in the field of education. This consideration was based on reaction to the common observation that the Hawthorne effect is most often introduced or posited as a causative variable in explaining significant results or the lack of them in innovative curricular and instructional studies. It was desired to conduct the research in a setting which would be highly replicative of existing research patterns.

Thirdly, it was felt that the experimental design should be adaptable to the possible role of indirect and direct cues as instigators of the Hawthorne effect. Indirect cues can be defined as experimental activities which inadvertently alert the subject to experimental participation. An example of such cueing would be unusual pretesting. Direct cues on the other hand would be obvious changes or activities which alert the subject to participation. Special directions or unusual physical changes are examples of such cues. To examine the possible influence of cues in creating awareness, a decision was made to create an experimental design which would incorporate both types or kinds of cues by introducing variables which could be considered as representative of each kind. The sections which follow describe the variables that were selected for incorporation into the field research design.

The major planning activities for field research are discussed separately. It should be understood that most of the procedures were carried out concurrently even though described separately. Specific sections which follow are (a) experimental design, (b) selection of experimental sample, (c) orientation of participating teachers, and (d) instrumentation and data collection.

Experimental Design

The original experimental design outlined in the project proposal was modified upon the advice of Dr. Julian Stanley who served as project consultant on design planning for the field research phase. Dr. Stanley consulted with the project staff early in the fall of 1962. This consultation was followed by a subsequent visit by the project staff to the Laboratory for

Experimental Design at the University of Wisconsin in January 1963 where the proposed experimental design was further critiqued. The critique sessions were tape recorded and recommended changes were reviewed and incorporated as deemed desirable by the project staff. The final experimental design resulted from the combined efforts of the project staff and the consultation activities described above.

Indirect Cues. In its elementary form, the comparative effectiveness of the investigation in education involves the presentation of a new and/or different method of instruction or curriculum (commonly called a treatment) to one group of subjects or schools and the resultant performance of such a treatment group is then compared to the performance of a control group not subjected to the treatment. Based upon the performance data obtained, an evaluation is made about the effectiveness of the two methods. A basic assumption is made that the only difference existing between the two groups during and at the end of the study is the method of instruction and/or curriculum introduced. It is at this point that the Hawthorne effect is often introduced to explain the superiority or lack of same in the experimental situation.

The development of the basic design gave cognizance to the traditional pattern of investigation by making an initial decision to employ as one indirect cue a new method of instruction and comparing it with a traditional, conventional, or on-going procedure. After consideration of possible alternatives, the new method selected was the curriculum materials developed by the School Mathematics Study Group (SMSG). It was felt that these materials constituted a sufficient departure from a traditional or conventional approach to mathematics instruction that they could be considered as being representative of new methods of instruction or curriculum. The SMSG materials were sufficiently developed at the time of the investigation that few difficulties would be encountered in providing necessary and appropriate materials for both students and teachers. In designating treatment combinations, the SMSG materials are subsequently referred to as the SMSG mathematics while the on-going, existing arithmetic program[1] is referred to as Conventional (Conv) mathematics.

The general hypothesis to be tested was that the introduction of these new and different curriculum materials in the area of mathematics might create among the participating students the feeling that they were in a special program and hence they would show greater mathematics achievement on appropriate instrumentation than students in the conventional program.

---

[1] The text for this program was Growth in Arithmetic published by Holt, Rinehart, and Winston in 1959.

In many experiments, it is not unusual to find that the introduction of a new method or curriculum is characterized by the creation of an unplanned change in the social relationships existing between the classroom teacher and participating students. Such a change could be a possible causative factor in producing unexpected achievement. An example of such a change from noraml classroom routine would be the presence of a different teacher brought in to teach the new method. It was decided, therefore, to explore this possible condition by introducing a change in the social structure of the participating classroom groups to serve as a second indirect cue. The specific procedure employed was to establish a teacher factor consisting of Regular and Different teachers for the class period devoted to mathematics instruction.

The Regular teacher condition consisted of the students being instructed by the teacher normally assigned to the participating classrooms used in the study. The Different teacher condition consisted of arranging for two teachers in the same building to exchange classrooms each day for the period of mathematics instruction. Before and after the mathematics instruction period, these teachers remained with their regular classroom. The hypothesis being tested was that exchange of teachers for mathematics would introduce a change in the classroom social relationships thus heightening the student's possible awareness that something special was going on in his environment and thus causing greater performance than where such a change was not made. It was also hypothesized that awareness of experimental participation and performance might be stronger in a classroom situation where both a new method and Different teacher were employed than in a conventional classroom with the Regular teacher.

A question might be raised as to whether or not the introduction of a Different teacher for the period of mathematics instruction who might possibly even be known to the students because of the need to carry out such an exchange in a single building would create a sufficient change in the social structure to effect the achievement results. If a school system already had separate instructors for special subjects (art, music, etc.) who enter a classroom for instruction, the appearance of a different instructor for arithmetic might not be a novel condition. On the other hand, students may be accustomed only to Different teachers for special subjects but not for a basic subject such as mathematics. To control for this possible factor, a decision was made that the experimental groups would have to be drawn from a school system which either did not have special subject instructors moving into classroom groups or where there was a history of not using special teachers.

Direct Cue. It has been noted by some writers on both educational research methodology and the analysis and interpretation of instructional and curriculum methodology, that students in some experiments have been told that they are in an experimental group and that such knowledge along causes students to do better than students not so informed. In most typical classroom experimentation, an attempt is usually made to avoid this type of a direct cue to students about experimental participation. Should awareness of experimental participation develop, an assumption is made that it was triggered by indirect cues, such as new textbooks, different tests, and similar materials and activities.

To test this observation and to study the possible role that such direct cueing might have, a decision was made to incorporate such a variable into the experimental design. This variable consisted of providing information to selected subjects at scheduled periods of time that they were in an experiment while some subjects were not told that they were in an experiment. In general, The Told condition consisted of informing students at the start of the experiment and at periodic intervals throughout the duration of the experiment that they were participating in an experiment on new mathematics materials. They were urged to do their best at all times and to participate to the best of their ability. Should students ask if they were in an experiment, the teacher was directed to answer positively. The Not Told condition consisted of providing no direct information to students that an experiment was going on. Under this condition, an attempt was made to consider the introduction of new mathematics material as a routine curriculum change. Teachers were warned against making any comments about experimental conditions. Should students ask if they were part of an experiment, the teacher was directed to reply negatively. Illustrative methods and patterns used to inform subjects about their participation in the experiment are included as Appendix A to this report.

Design Outline. The final experimental design for the field research consisted of three treatment factors, each having two levels. This design permitted a study not only of the main effects but also the interactions of the various factors. For purposes of analysis, all factors were considered to be random factors.

An outline of the final treatment combinations[2] is presented as Figure 1. The several conditions represent a continuum of combinations such that Different Teacher-SMSG-Told would represent the maximum combination of indirect and direct cues introduced into an experimental classroom while treatment combination of Regular Teacher-Conventional Math-Not Told would represent a control group situation. Under a general hypothesis that awareness of experimental participation is created by such direct and indirect cues, differences in achievement should be at a maximum between these two treatment combinations if the factors employed do in fact magnify novelty, enthusiasm, attention, and similar subject reactions which are thought to cause awareness and thereby increasing student motivation, under experimental conditions, to achieve higher levels of achievement.

Selection of Experimental Sample.

The selection of the sample had to be undertaken within several constraints imposed by the experimental design. The decision to use SMSG mathematics materials imposed a constraint on the sample in terms of the level of the education at which the research could be conducted since the materials were only developed sufficiently well enough to be used at the elementary grades. A further constraint was that the students involved needed to possess sufficient verbal abilities so as to be able to respond to written achievement tests and possible oral interviews. Because some students would be told and others not, some geographical isolation of students from each other also would be necessary. Further, it would not be possible to consider the student as the sampling unit since the introduction of the new method and different teacher would require that classroom groups be utilized. Using the latter as a sampling unit, at least two classrooms for each treatment combination would have to be involved since the participating teachers under the Different Teacher factor would have to change classrooms. It was originally proposed that teachers move from school to school in order to meet this condition. Consideration of this idea, however, led to it being considered as too involved and impractical.

---

[2] The several treatment combinations have been abbreviated as indicated in Figure 1 in order to avoid unnecessary long descriptions. The reader should become familiar with this notation since it is used in the summary tables of the analysis of results.

Figure 1.—Experimental Design Variables and Treatment Combinations

Indirect Cues		Direct Cue	Abbreviation
Method	Teacher		
Conventional Mathematics	Regular Teacher	Not Told	C-R-NT
		Told	C-R-T
	Different Teacher	Not Told	C-D-NT
		Told	C-D-T
SMSG Mathematics	Regular Teacher	Not Told	S-R-NT
		Told	S-R-T
	Different Teacher	Not Told	S-D-NT
		Told	S-D-T

The decision regarding the final grade level to be used also considered that the introduction of new materials should occur at a time in the typical curriculum sequence when natural transitions are made. One such transition occurs between third and fourth grade. Consequently, a decision was made to initiate the experimental design at the fourth grade level. It was felt that initiating the field research at this level would satisfy most of the constraints noted above.

Since one objective was to study the effect of the duration of an experiment as a possible control for the Hawthorne effect, the use of fourth grade initially would permit the continuation of the study into a second year without too much difficulty for teachers, students, and project staff.

Under the experimental design established, the basic sampling requirements called for sixteen classroom groups in order that these would be sufficient degrees of freedom for the statistical analysis. Two classrooms were to be assigned to each of the eight treatment combinations within one treatment combination assigned to each school. Thus, each treatment combination would have an n=2.

Superintendents of local school systems which might supply the needed number of classrooms were contacted for possible participation in the field research. The general nature of the project and the specific purpose of the field research were explained. Information was also provided about the teacher training program to be installed as part of the project. This action was considered important to take at this time since knowledge that a participating teacher had of the new mathematics subject matter might adversely affect the results of the study. It was explained that a professional mathematics education person was to be added to the project staff for the purpose of conducting the in-service training for both the new method or SMSG teachers and the conventional or old method teachers. It was anticipated that the in-service program would provide reassurance to teachers as they undertook the task of becoming familiar with the new mathematics.

The Columbus Public School System consented to participate in the project because responsible persons felt that the project itself was worthy of support plus the fact that benefits would be gained in securing additional elementary teachers trained in the new mathematics under careful supervision. The system imposed one condition upon their participation which related to the desire to maintain geographical isolation of treatments.

The Columbus System retained the right to select the schools which could be used but with a proviso that the final selection of schools would be as relatively homogeneous as possible in terms of socio-economic location within the city, teaching personnel, and student ability and achievement levels. A further agreement was reached between the project staff and the participating system that any schools selected should be ones where there had been a minimum of prior experimentation plus assurance that additional experimentation would not go on during the two year duration of the project. It was further agreed that the participating classrooms would be withdrawn from the student teaching pool.

Early in the spring of 1963, the project director met with a group of elementary principals and teachers from possible participating schools selected by the system to orient them to the nature of the project. From this meeting indications of willingness to participate were obtained from the principals. The responsible officials in the Columbus System then selected the final eight schools for participation within the agreements noted above.

A second meeting was held with the principals and potential participating teachers to give them a complete description of the experimental variables and conditions which would be randomly assigned to each participating school. At this point in time no participating teacher or principal was aware which treatment condition would fall to their school, yet there were expressions of complete cooperation.

The final selection of teachers to be involved with the study remained at all times with the principal of each participating school. Criteria suggested for final teacher selection included the following: at least three years of teaching experience, willingness to participate in bi-weekly meetings, and willingness to teach a new method of mathematics--new and different from that which they had been teaching.

Orientation of Participating Teachers.

As noted above, a professional mathematics education person was added to the staff of the project in June of 1963. This Associate in Mathematics, as he was designated, undertook to prepare necessary materials to be sent to the participating teachers during the summer and for use in the forthcoming in-service sessions.

During the summer of 1963, the treatment conditions were randomly assigned to the classroom groups selected from the participating schools. A teacher's edition of the fourth grade SMSG textbook was sent to each teacher requiring one along with an information bulletin outlining the treatment condition for that school. In those schools where the subjects were not to be told that they were in an experiment, emphasis was given to this point in the bulletin and clarified in later meetings with the teachers. The teachers were informed that it would be necessary for them to *deny* to the students, and possibly even to the student's parents and to fellow teachers in the same building, that they were involved in an experimental situation.

The textbook for the SMSG mathematics was completely different from the conventional arithmetic textbooks being used in that it was yellow in color and paperback. Before being issued to the classrooms, the books were stamped with Columbus Public School identification and then delivered to participating schools in as normal a manner as possible.

Two meetings for all teachers were held before school began in the fall of 1963. The SMSG and Conventional teacher groups met separately. Each group of teachers was re-oriented to the treatment combination which had been randomly assigned to their school during the early part of the summer. Although the SMSG teachers had been provided a Teacher's Edition of the SMSG textbook, these meetings served to acquaint each teacher further about the overall program and particularly with the material and concepts to be taught during the first few weeks of school. At these pre-school meetings, a schedule was worked out which enabled the Associate in Mathematics to observe in the classrooms and to conduct an in-service program in the same school on that day. The visitations were made to each school on a pre-planned day once each two weeks during the entire two years of the study. The scheduled in-service and observation periods permitted the Associate in Mathematics to make sure that the conditions of the treatment in that school were being observed. The observation scheduled also permitted an attempt to equate for all groups the attention or novelty that the presence of such a visitor might cause.

The decision to run two concurrent in-service programs for the teachers was based on research done by Rudell and Brown (70) and Houston and DeVault (35) which implied that conducting an in-service program for one group of teachers without a comparable program for the other group of teachers contaminated inherent differences between the methods themselves.

At each participating school, an in-service program was prepared which kept pace with the concepts and progress of both classes. Adjustments from school to school allowed each program to be tailored to individual needs and achievement levels. During the arithmetic period, the Associate would visit in only one classroom alternating his visits between the two classes at each school. Teaching techniques, content presentation, and the manner in which the subjects received instruction were observed. The observations made at this visitation served as a core part of the in-service conferences normally held after regular school hours.

Experimental Log. The Associate in Mathematics maintained an experimental log of his daily visits for the two-year period of the study. This log was developed to enable the Associate to record in a systematic and regular manner any conversations or incidents which could lead either a teacher, principal, or the Associate to believe that a subject was expressing an increasing or undue awareness of his experimental environment. For example, the Associate entered a room in the Regular-Conventional-Not Told school after hours for the conduct of an in-service meeting during the second year of the study. He was greeted by a fifth grade subject with the remark, "Oh!, you're the man who came last year and taught our teachers arithmetic." An observation entry recorded involved a fourth grade subject who suddenly stood up in class and said, "Mrs. ________, is this experiment we're in a success yet?". Another entry related to concept formation and involved a fourth grade male who had finally learned not to rewrite a peculiar problem. The subject, when called on to read the answer to a problem replied, "Four minus seven has no answer, but if it did it would be minus three!" In retrospect, the maintenance of the experimental log should have been expanded more formally to include each teacher as will be indicated in the section of the report dealing with the post-treatment interview of the teachers.

Teacher-Student Continuity. A major question with which the project staff had to deal toward the end of the first year of the study was whether or not the participating classrooms should be set up so that both teachers and students in the several treatment conditions remained together for the second year of the study. Employment of a principle of continuity would require that fourth grade teachers be asked to move with their students up to the fifth grade. As for the students, the maintenance of intact classroom groups from the fourth to the fifth grade level, with the same teachers, would be a departure from normal school routine. Assistance on this decision was secured from the Elementary Supervisor for the Columbus Public School System toward the end of the first year.

Teacher personnel turnover data were examined to see what information might be provided from this source. It was noted that in five of the eight schools, one teacher had left the school. In another school, two of the teachers were departing. Three of the teachers had departed temporarily for family reasons. Two teachers were leaving the system and two others were being shifted to supervisory or counseling positions. If the decision was made to continue during the second year of the study with the first-year teachers, the turnover rate would have been about 35 percent. Discussions with the teachers and the elementary supervisor highlighted the fact that the teachers did not want to move up a grade because this would require them to become oriented to and develop new materials for all fields and not just mathematics. For these reasons, a decision was made to continue the pattern that the students would normally follow with regard to teachers in moving from fourth to fifth grade. Thus each student would be in a self-contained classroom for the second year of the study with a new teacher for the fifth grade.

A similar decision was made with regard to the students as they moved from the fourth grade to the fifth grade. In those schools where there were more than two classrooms, a decision had to be made whether or not to intermix only the two experimental classrooms or all classrooms in the building. The latter procedure would have had the effect of reducing the total number of subjects since placement in a non-participating room would eliminate some subjects from the study. After consultation with the principals involved, it was decided that any shifting of students between classrooms would be done within the normal procedure of the school. An agreement was reached, however, that any intermixing of students from classroom to classroom would be done within the two experimental rooms in so far as possible so as to maintain the class size and not contaminate the experimental classroom with new students unless they transferred into the school. Any such transfer students would be accepted into the classrooms as would any new student in a school where there were only two classrooms.

Instrumentation and Data Collection.

The section on the experimental design for the field research phase pointed out that the primary function of this activity was to set up a situation comparable to many typical methods experiments with the addition of selected factors or variables which were thought might produce differential student achievement levels among the various factorial combinations. It was hypothesized that students who received the different factorial combinations would vary also in the degree of their awareness of participation in an experiment. This hypothesis was established to test the commonly held assumption that students aware of experimental arrangements, particularly new or novel ones, generate a feeling of special attention. If this assumption is true, one might be lead to expect that the more aware students would have greater achievement than less aware students.

Recognition is given to the fact that there would be some difficulty in partialing out that portion of total achievement due to instructional and content processes and that due to awareness so that it could be stated empirically how much of the total achievement was due to feelings of awareness and their subsequent motivational effects upon students. To test the hypothesis, the approach used in this study was to secure some measure of both achievement and awareness. The purpose of this section is to describe the instruments used and/or developed to secure quantitative measures of these two variables.

Arithmetic Achievement. The basic approach to the measurement of mathematics achievement was that of securing gains in achievement over a two year span of time. The general procedure was that of traditional pretest and posttest differences. Consideration was given to the problems of measuring gains in view of recent discussions of this problem in educational research (30). On the basis of these discussions, a decision was made to use the "true" gain procedures as outlined by Lord (47, 48).

Because of the unique nature of the SMSG mathematics, any instrumentation utilized to secure such gains would have to be a type that could not penalize either group. After careful consideration of possible alternatives plus securing recommendations from other persons who had conducted studies in arithmetic achievement that the completely verbal presentation of problems not normally found on most standardized arithmetic achievement tests had been shown not to favor a particular teaching method (56), a decision was made to use the Sequential Tests of Educational Progress: Mathematics, Level 4 (Grades 4, 5, and 6). This test had equivalent forms available which facilitated repeated measurements over the duration of the project. Since the validity and reliability of this test are reported in the

Technical Manual for the STEP tests, such information is not repeated here (73).

The STEP test was first administered in the fall of the fourth grade and again in the spring with a similar pattern being followed for the fifth grade. Since this test was not part of the regular Columbus School System testing program, its introduction might have had some impact on students but no evidence was gathered on this particular point.

Awareness Measurement. The problem of measuring awareness of experimental participation was complicated by two major considerations. First, the lack of any prior or existing instrumentation and second, the relatively large number of students involved in the study. Because of the somewhat unique nature of this assessment, the development of the awareness instrument is described in detail in Appendix B. Consideration was given to using an oral interview technique but abandoned in favor of a paper-pencil inventory which would produce a quantifiable measure of awareness. The final instrument was a Likert-type scale which could be administered by the teachers in the normal classroom situation.

The awareness instrument was administered at the several participating schools within a one week period after the completion of the final administration of the STEP Mathematics test at the end of the fifth grade.

Supplementary Measures. The arrangements made to conduct the study within one administrative system made it possible to collect additional data from the classrooms participating in the study as a natural and routine part of the regular Columbus Public Schools testing program. In addition to the achievement and awareness measures described above, data from the instruments described below administered by the participating system were obtained and made part of the individual student record.

1. Columbus Arithmetic Test. This test is rewritten each year by the supervisory staff of the Columbus System and consists of non-verbal problems selected on an equal proportion basis from the third and fourth grade conventional arithmetic texts. The test is normally administered as a diagnostic test in the fourth grade.

2. Metropolitan Achievement Test, Elementary Reading, Grades 3 and 4. This test is administered early in the fall to fourth grade pupils. A Word Knowledge and Comprehension scores are obtained for each student.

In this study, the test was administered before any treatment conditions were introduced to the participating schools.

3. Kuhlman-Anderson Intelligence Tests. This test was administered to each student during the second grade. Scores for students participating in the present study were secured from student records. Due to the turnover of pupils in some of the participating schools, it was not possible to secure complete data fro all students.

4. California Test of Mental Maturity, Short Form, Elementary Grades 4 thru 8, Form S. This test had been regularly administered during the spring of the fifth grade but was shifted system-wide during the project to the first week of testing in the fall of the fifth grade. In addition to total scores, subtest scores were recorded for each student in the project.

5. Stanford Achievement Test, Arithmetic and Reading. This test was new to the system during the course of the project. The Reading test provides scores on Word Meaning and Paragraph Meaning. The Arithmetic Test includes sections on Arithmetic Computation, Concepts, and Applications.

Administrative Arrangements for Testing. The testing schedule for the several tests noted above appears as Table 1. Recognizing that the administration of the STEP tests to only selected classes in the participating school might create some unusual attention in and of itself, a decision was made to test all students at the grade level in which the study was being conducted. In some schools, only the two classrooms involved in the project were tested. In those schools containing more than two classrooms, all other classroom groups were tested at the same time. Under this procedure, the data collected from non-participating classroom groups could be used to provide for an external control group, or what some methodologists refer to as a control-control group.

The data on each student from the several test administrations along with other data thought to be relevant was recorded on forms which served as input documents for latter data processing.

After the administration of the STEP test in the fall of 1963, a technical problem was encountered with regard to the scores that should be used for the subsequent data analysis.

Table 1.--Instrumentation and Data Collection Schedule[3]

Date* Week	Instrument Administered
March 15, 1961 (Grade 2)	Kuhlman Anderson Intelligence Test
September 10, 1963	*STEP Mathematics (I)
September 17, 1963	Columbus Arithmetic Survey
October 15, 1963	Metropolitan Reading Test
May 26, 1964	*STEP Mathematics (II)
September 8, 1964	*STEP Mathematics (III)
September 15, 1964	California Test of Mental Maturity
March 9, 1965	Stanford Achievement - Reading
March 16, 1965	Stanford Achievement - Arithmetic
May 17, 1965	*STEP Mathematics (IV)
May 25, 1965	*Awareness Scale
June 1, 1965	*Teacher Interviews

Notes: 1. Individual schools varied with regard to specific testing dates so only week of testing has been indicated.

2. Tests marked with asterisk were introduced for purpose of the study.

---

* [3] The STEP Mathematics test was administered four times during the two years of the project. It was administered as a pretest in September 1963, again in the spring of 1964, the fall of 1964 and in the spring of 1965. The purpose of conducting four administrations, with a possible cueing factor due to repeated tests, was to permit an analysis of data at the end of one year and two years period in an attempt to secure information about the short versus long range duration of studies and the diminution of the Hawthorne effect. For purpose of convenience, these administrations are identified by Roman numerals. Thus, the test in fall 1963 is STEP I and the administration in spring 1965 is STEP IV.

The general nature of this problem was that raw scores of 13 and below on the test are assigned the same standardized scale score. For some of the classes, there were many scores below 13. Hence, if standardized converted scores were used, they would have the effect of reducing the range of scores since any raw score below 13 received the same scaled score.

Since the pretest means for 5 of the 16 classrooms were at 13 or less, or at the lowest converted score level, a decision had to be made as to whether or not the raw scores or converted scores should be used in the analysis since the results might be effected by the particular data used. A decision was made to conduct statistical analyses of both converted and raw score data but only the raw score data is presented in this report since comparison of results from the two analyses were quite similar.

One of the principal hypotheses of the study dealt with a concern over the gain in mathematics achievement and its possible relation to awareness at the end of the two year period. The analysis of data only from the two year period, however, would not permit an exploration of the hypothesis relating to duration of the study and possible achievement gains. To test this hypothesis, a decision was made to conduct separate analyses at the end of the first year, again at the end of the second, plus the full two years. It was hoped that by conducting analyses at the end of these points in time, the effects of the indirect and direct cues could be studied to determine their influence during the first year and second years respectively as well as examining their influence over the full two year period. Ideally, it might have been desirable to include measures of awareness at the end of the first year but a decision was made not to include any such measure since it was felt that its administration might provide a cue to the students thus effecting the results at the end of the full two years.

Analyses at the end of each year period had to consider the problem of subject attrition. Many students were lost from the experimental groups over the two year period for a variety of reasons, including failure to complete all relevant measures. Consequently, the analyses were made at each time period using those students who had the necessary data at the end of each of the several time periods.

Teacher Interviews. In many experimental education situations, the classroom teacher becomes an experimenter surrogate as described by Rosenthal (68) or may even at times assume the role of a subject. In most educational studies, the teacher serves as an intervening variable between the experimenter and his subjects--the classroom students. The inclusion of the classroom teacher as an active participant makes her an unpaid member of the project staff which can have an unknown influence on her activities. Further, her participation may be affected if she is made to feel like an observed subject. On the other hand, the teacher can serve as a valuable source of data with regard to the impact of the treatment conditions on student learning and reactions. Accordingly, arrangements were made to interview the participating teachers at the end of the study.

The principal function of the interviews was to discover some of the unexposed ideas or incidents concerning the Hawthorne effect as it was defined for purposes of this study. A set of questions was developed to serve as a basis for a structured interview situation. The interviews were to be conducted after the collection of all student data and before the termination of the second year of the study. The outline of the structured interview is presented as Table 2.

Table 2.--Teacher Interview Questions

A. Questions Centering on Teacher Observation in School Situations and Atmosphere.

   1. What appeared to be the general school atmosphere regarding participation in the experiment?

   2. Was there any general positive or negative attitude expressed regarding the experimental conditions assigned to the school?

   3. Does your school have any special class conditions (homogeneous group) which might have effected the results?

   4. What parental reactions, if any, developed regarding the experimental conditions?

B. Questions Centering on Teacher Observation of Student Reactions and Attitudes.

   1. Do you think students were generally aware of their participation in an experiment?

Table 2.--Teacher Interview Questions (Continued)

2. If students seemed to sense that they were in an experiment what possible direct or indirect cues do you think were picked up which would give them this information? (Follow up with prompt about independent variables).

3. Do you feel the students were better motivated due to their special treatment than would be the case for the normal or more typical classes you had experienced?

4. Do you think students were negatively or positively motivated by participation in the experiment?

C. Questions Centering on Teacher Reactions.

1. What previous experience have you had with regard to participating in educational experiments?

2. How did you accept the fact that you were participating in an experiment? Was your motivation different than if you had not participated?

3. Did complying with the experimental conditions for your school impose any hardships? If yes, how did it affect your participation in the experiment?

4. Have you at any time discussed the experimental conditions and measures with other colleagues?

5. Do you feel you worked harder knowing that the results of the experiment would be reported to the Superintendent of Schools as well as on a national basis?

The interviews were conducted by the principal investigator and the Associate in Mathematics at each individual school with all available participating teachers from the fourth and fifth grade present. In most cases, this interview situation was the first time that many teachers had seen or had the opportunity to talk with the principal investigator since the initial orientation periods. This action on the part of the principal investigator was deliberate and taken on the desire to reduce as much as possible the influence of external persons to the experimental situation. It was felt that the continuous or frequent presence of such a person in the participating schools would have a negative influence on the experiment by perhaps highlighting the situation. The question of how much and in what manner the presence or absence of the principal investigator in an experimental situation such as the one conducted in this project might well be the subject of further study. In the present study, perhaps he should have appeared more often in selected classrooms in order to heighten the desired effect.

As noted, the interviews were held directly in the schools and tape recorded except for one participating school. The teachers at this school took the position that they did not want their remarks recorded nor did they particularly want individual type interviews. Consequently, the principal investigator and Associate in Mathematics invited these teachers to a luncheon at which time an informal conversation was held but directed using the structured interview questions as a guide to the conversation. Of the thirty-two teachers who had worked with the participating classrooms, only twenty-four or 75 percent were available for the interviews.

II - Literature Analysis Methodology

The basic purpose of the literature analysis phase of the total project effort was to determine if useful information regarding the possible nature and role of the Hawthorne effect in educational research could be obtained by conducting a systematic examination of completed research studies comparing experimental to conventional methods as reported in educational literature. As such, it was a separate and distinct activity from the field research phase.

Specifically, the literature analysis phase was directed to questions 1, 2 and 4 of general question A. It was also anticipated that useful information might be obtained which would help to answer the questions under general questions B and C provided that sufficient detail could be secured from the published report.

The purpose of this section is to describe the procedures employed in this phase of the total project. For purposes of presentation, the section has been subdivided into parts dealing with (a) Research Areas Investigated, and (b) Abstracting Procedures.

Research Areas Investigated

The selection and definition of the specific areas of completed research on comparative experiments to be examined was initiated by using the 1960 edition of the Encyclopedia of

Educational Research (29) as a primary reference and guide. Nearly all of the titles of the major areas finally selected, the definitions of these major areas, the subgroups identified within these areas, and the levels of education at which research was done in these areas were directly derived from this source.

Some additions and changes to the original list of research areas derived from the above source were made by the project staff in order to better meet the specific demands of the projected analyses. The final major areas and subgroups for use in the literature analysis phase were chosen on the basis of what seemed logically relevant and then subsequently modified by observation of the kinds of studies abstracted in the early stages of the analysis.

Twenty-six major areas of completed research were finally selected. These areas were then combined into three primary categories for the purpose of clarifying the different types of studies that seemed most valuable to examine. The major categories were established on the basis of the similarities among the experimental variables manipulated in the studies found in the major areas of each specific category. Thus, the criterion for the classification of a major area to one of the four categories was the general type of the experimental variable manipulated in the studies within that major area. The extensive use of the experimental variable as the fundamental criterion for the classification of studies was based upon the observation that the description of the experimental variable in the research report served best to define the nature and focus of interest for the experiment in question. It was felt that on the basis of the experimental variable alone, a sufficiently accurate classification of studies could be made which would facilitate the final analysis stage of the literature phase. A fourth category was established which included relevant materials but which could not be classified as research studies. The four principal categories along with a short description of each in terms of the types of experimental variables that were commonly manipulated in the studies found in each category are presented below.

Subject Matter Areas. Included in this category were those studies employing specific teaching methods as experimental variables which were particularly applicable to individual subject matter areas. For example, in the major area of Social Sciences and subgroup of American History, the experimental variable (new teaching method) that might be used was the "discussion" method. The new teaching method used as an experimental variable was usually dependent upon the content of the specific subject matter; the specific new teaching methods were often peculiar to one major area only.

Facilitative Devices or Procedures for Improving Teaching. Included in this group were those types of studies which employed special classroom devices or types of school organization, (e.g. teaching machines, homogeneous grouping, etc.) as experimental variables (new teaching method). Usually a subject matter was involved in these studies, but only to provide a vehicle or context in which the experimental variable was manipulated. Consequently, the same new teaching method may be used in experiments in different subject matter.

General Teaching and Learning Methods. This category was somewhat of a catch-all for those major areas which had little in common with the major areas in the other categories. However, the major areas here did have in common the fact that they all employed experimental variables not directly related to subject matter areas *per se* or to special devices or procedures for improving teaching.

Hawthorne effect Special Interest Areas. This category was the only exception to the rule of using the experimental variable as a classification criterion. The major concern of interest here was those studies, articles, or reports that directly or indirectly dealt with discussions relevant to or about the Hawthorne effect, its further definition, and the conditions necessary for the production and/or control of the effect.

The four principal categories along with the twenty-six major areas of research and the subgroups related to each major area are presented as Part I of Appendix C. Following the title of each major area are the levels of education at which the research studies within a major area have usually been conducted using a special notation for each educational level as given at the end of the complete list of major areas. The use made of the major areas, subgroups, and educational levels in abstracting, filing, and final analysis is outlined below.

Abstracting Procedure.

When examining a research report to determine its potential contribution to the literature analysis, two primary criteria were used to determine the relevancy of a study. These criteria were that the study must have an experimental design comparing new versus on-going teaching methods, techniques, or procedures, and employ an experimental design in which an experimental group received a new method and a comparison or control group received an on-going method.

With regard to the first criterion, a new method was defined as a method which is perceived by the experimenter as a novel, different, or unique method of instruction, classroom or school organization, and so on. This definition assumed that the population involved had not been exposed to any such method previously. An on-going method was defined as a condition which was perceived by the experimenter as one with which subjects had previous experience, either directly having been taught by the method or indirectly having knowledge that such a method was used to teach other students in situations similar to the present one. Because the new and on-going methods are defined in terms of the particular experimental situation of each study, new and on-going are seen as relative, not absolute, terms. Thus, what may be a new method in one study may be an on-going method in another study. The two terms, new and on-going, were used by the project staff to describe the use made of particular methods in experimental teaching methods studies. They are not referred to as such by investigators in their reports of research. Consequently, the abstractor identified the new and on-going methods by inference from the description of the teaching methods used in the study. This process was usually less difficult than it may seem because of the conditions set by the second criterion.

With regard to the second criterion, an experimental group was defined as one receiving the new method. This new method was usually referred to as the experimental variable by the investigator. A comparison group was defined as one which received an on-going method. Individual investigators may often call a group which receives an on-going method "control" group. In accepted research terminology, this usage of the term control with the on-going method is incorrect. A true control group receives no treatment conditions. Other investigators may term as an experimental group, one which receives (in the project's definition) an on-going method. In this project's terminology, and for its purposes, this experimental group is actually a comparison group and was treated as such by the abstractor.

If a study examined for possible use failed to meet either or both of the two primary criteria described above, it was not subsequently abstracted. It was recorded, however, as having been examined.

Two different types of abstract procedures were employed. The first and most commonly used form involved the abstracting of the essential details of experimental teaching methods studies for the ultimate purpose of comparison of such studies in the final stage of the literature analysis phase. The other abstract form involved references giving direct mention of the Hawthorne effect and those experimental design references which had implications for analysis because of their relation to the creation or control of the Hawthorne effect. Because of the proposed comparison of experimental teaching methods studies in the final stage of the literature analysis, it was necessary that a standard abstracting format be used to facilitate effective comparisons. Likewise, the language used in abstracting research studies was standardized as much as possible so that there would be minimum confusion over meanings and implications.

The standard abstracting format for the experimental teaching methods studies is presented as Part 2 of Appendix C. This format should convey the types of information collected for the comparison and final analysis of studies. In order that the process of abstracting be more explicitly known, since it was a data collection procedure, the specific procedural steps followed are presented as Part 3 of Appendix C.

The principal sources for the selection of research studies to be included were generally recognized professional periodicals in the field of education. An examination of all the possible periodicals and journals which have published research studies relevant to the purposes of the study would have been an almost impossible undertaking. The set of periodicals and journals finally chosen to be reviewed on a systematic basis were determined after an examination of possible periodicals which regularly reported the results of experimental methods studies. Further, it was decided to limit the examination of the final set of journals to the general time period from the years 1953 to 1963.[4] The time period cited was chosen primarily on the idea that if researchers were concerned about the Hawthorne effect in their projects, efforts to control for it would more likely appear during this period of time than might be the case in studies conducted much earlier when the phenomenon was not so widely publicized. The time limits also served to make the review process more manageable in terms of the work load involved.

---

[4] It should be noted that considerable latitude was exercised in varying from these time limits because of leads given in one study to a research activity carried out at another time period.

In addition to selected periodicals in education, efforts were made to secure research studies from other sources such as final reports of projects funded by the U. S. Office of Education, private foundations, instructional research conducted by the military and similar agencies, dissertation abstracts, colleges and universities. In addition to the above sources, arrangements were made to secure reports of available abstracted research from such operations from the Center for Documentation at Western Reserve University, Science Information Exchange, Defense Documentation Center, and individuals who were known to have made collections of research studies for other purposes.

From the above sources, approximately 350 abstracts of research studies were secured which were considered to be useful to the investigation within the general time period noted and the time available to do abstracting. This set of abstracts served as the basic data for the subsequent analyses conducted during this phase of the project. Recognition is given here to the limitations of the above procedure in terms of the final sample studies and the abstracting process itself and their possible influence on the analyses subsequently conducted.

Chapter III

Field Research Results

Introduction

The purpose of this chapter is to present results from the field research phase of the total project. The results which follow have been separated into two sections. The first section presents those results with regard to questions 3 and 4 of general question A, questions 1 and 2 of general question B, and question 1 of general question C. The essential nature of these questions related (a) to the feasibility of establishing an experimental classroom instructional situation so as to determine the role that experimentally introduced direct and indirect cues might have on subsequent achievement and (b) to investigate the possible relationship between achievement and awareness of participation in the experimental situation. The second section presents results with regard to question 4 of general question A and question 3 of general question C. These questions were concerned primarily with the desirability of extended duration of an investigation as a control procedure for the Hawthorne effect. The sections titled Cues, Awareness, and Achievement and Duration of Study present findings from the field research which relate to the above questions. This chapter also presents the results from the interviews of the participating teachers conducted upon the completion of the field research phase.

Cues, Awareness, and Achievement

As noted in the chapter on methodology, an experimental situation was created to determine if variables or factors identified as direct and indirect cues would have any effect on mathematics achievement in a classroom situation. The description of the methodology employed for the field research as presented in the same chapter has indicated that several pretests were administered to the participating classroom groups. One test, the STEP Mathematics test, was not part of the regular school testing program but was introduced to provide a criterion measure for subsequent data analysis.

Analysis of Initial Measures. The first analysis was conducted to determine if the several classroom groups under the eight treatment combinations differed by initial performance on the criterion test and other measures. Table 3 presents for each class within the several treatment combinations the pretest raw score means and standard deviations together with the

Table 3.--Raw Score Means and Standard Deviations for Selected Pretests

TREAT-MENT ABBRE-VIATION	CLASS	STEP MATHEMATICS			METROPOLITAN						COLUMBUS ARITHMETIC			K.-A. IQ (2nd GRADE)		
					WORD KNOWLEDGE			COMPREHENSION								
		n	$\overline{X}$	s	n	$\overline{X}$	s	n	$\overline{X}$	s	n	$\overline{X}$	s	n	$\overline{X}$	s
C-R-NT	1	28	17.3	4.7	29	29.6	10.8	29	24.8	8.7	29	37.38	7.7	31	101.4	8.5
	2	32	17.2	5.4	34	28.7	11.1	34	24.9	7.7	34	36.65	7.9	30	101.4	9.2
C-R-T	1	25	17.0	4.5	21	30.1	9.9	21	25.9	7.5	29	37.38	8.2	10	95.9	12.6
	2	26	17.7	6.2	24	33.1	10.0	24	27.3	8.3	25	38.76	8.8	22	103.9	7.6
C-D-NT	1	33	11.8	4.2	34	24.4	8.9	34	20.9	7.0	32	32.16	8.3	30	96.97	10.1
	2	32	14.5	5.8	34	24.2	8.2	34	21.6	7.3	33	32.85	7.6	30	100.3	7.4
C-D-T	1	32	16.4	4.7	32	24.7	8.7	32	24.1	5.9	32	34.28	7.7	30	101.0	5.9
	2	26	11.27	4.8	27	18.9	5.5	27	18.3	5.2	25	25.52	6.9	23	96.2	8.0
S-R-NT	1	40	12.4	5.5	38	23.5	10.8	38	18.6	8.0	39	29.69	8.7	34	98.2	9.6
	2	37	13.4	5.3	37	21.7	9.2	37	18.5	7.6	-*	---	---	35	93.5	8.9
S-R-T	1	32	12.1	4.4	32	25.1	9.0	32	21.2	7.2	30	32.17	7.3	25	98.2	7.8
	2	32	14.7	6.6	30	24.6	11.8	30	21.0	8.2	29	32.41	8.9	26	95.5	11.1
S-D-NT	1	30	18.4	6.1	31	31.9	12.0	31	28.1	9.3	30	37.43	9.2	29	100.3	9.5
	2	29	16.1	6.0	30	31.5	11.4	30	28.3	7.7	30	34.8	8.5	27	102.6	7.6
S-D-T	1	29	18.2	8.7	31	33.6	11.5	31	30.4	8.4	31	41.45	7.7	22	98.9	6.4
	2	29	20.2	7.4	29	34.5	9.8	29	30.0	7.8	29	36.79	10.1	26	102.6	7.2

*One teacher returned the Columbus Arithmetic Tests for student use without recording any scores.

number of subjects on which the descriptive statistics were calculated. Since the basic sampling unit was considered to be the class, the mean of the two means for the pretests from the two classes under each treatment combination were calculated and are presented in Table 4.

Table 4.--Treatment Means for Selected Pretests

TREAT-MENT	n	STEP	METROPOLITAN		COLUMBUS ARITHMETIC	K.-A. IQ (2nd GRADE)
			W.K.	COMP.		
C-R-NT	2	17.2	29.2	24.8	37.0	101.38
C-R-T	2	17.3	31.6	26.6	38.1	99.9
C-D-NT	2	13.1	24.3	21.3	32.5	98.6
C-D-T	2	13.8	21.8	21.2	29.9	98.6
S-R-NT	2	12.9	22.6	18.6	29.7*	95.8
S-R-T	2	13.4	24.9	21.1	32.3	96.9
S-D-NT	2	17.3	31.7	28.2	36.1	101.4
S-D-T	2	19.2	34.0	30.2	39.1	100.7
TOTAL	16	15.5	27.5	24.0	34.6	99.2

*One teacher returned the Columbus Arithmetic Tests for student use without recording any scores.

Inspection of Table 3 reveals that the means of the several classes under each treatment condition were quite dissimilar on the STEP test as well as on the other pretests. The heterogeneity of initial performance is also highlighted in Table 4 which shows that the initial means for each treatment combination were highly variant, particularly for the criterion measure. Although the treatment combinations were assigned to schools on a random basis, it appeared that the procedure used by the participating school system to select schools with the assurance that some degree of homogeneity of classes would be maintained actually resulted in a group of heterogeneous classes being selected to which treatments were subsequently assigned.[1]

An analysis of variance conducted to determine if the observed initial differences between the treatment groups on the criterion STEP test were statistically significant is presented in Table 5. Tests of the main effects were not significant as was true for two of the interactions and the third order interaction. The Teacher-Method interaction, however, was significant. A graphical analysis of this interaction is presented in Figure 2 where it can be seen that the classes receiving the SMSG Mathematics with a Different teacher and the classes receiving Conventional Mathematics with the Regular teacher had higher initial scores than did the classes receiving SMSG Mathematics under the Regular teacher or Conventional Mathematics with a Different teacher. In view

Table 5.--Summary Table for Analysis of Variance for STEP Pretest Means

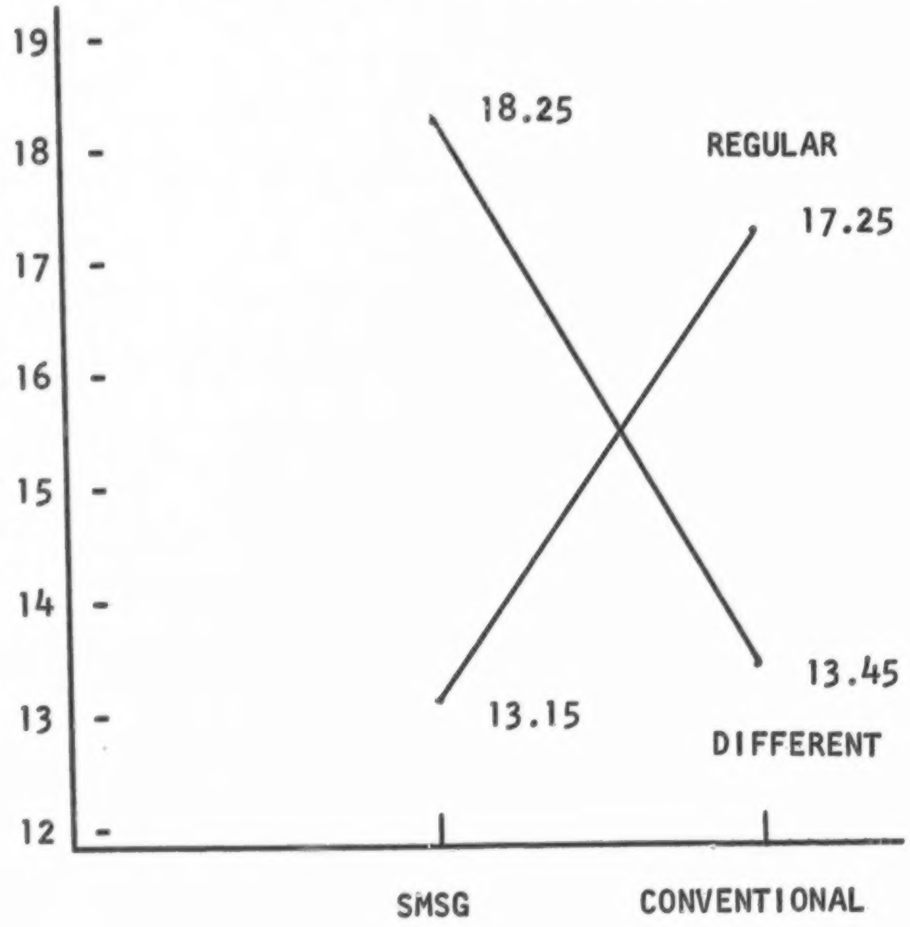
SOURCE	SUM OF SQUARES	df	MEAN SQUARE	F
Teacher	1.1692	1	1.6192	0.438
Method	0.3630	1	0.3630	0.982
Informing	2.5360	1	2.5360	0.6863
Teacher-Method	78.3668	1	78.3668	21.2083**
Teacher-Informing	1.0869	1	1.0869	0.2941
Informing-Method	0.6602	1	0.6602	0.1787
Teacher-Method-Informing	0.1870	1	0.1870	0.0506
Error	25.8659	7	3.6951	
TOTAL	110.6850	14	---	---

$F(1,7;.05) = 5.59$ $F(1,7;.01) = 12.25$

---

[1] In retrospect, it might have been more desirable to secure sufficient information on the schools so as to avoid this heterogeneity or possibly conduct an examination of the pretest performance before assignment of treatments.

Figure 2.--Teacher-Method Interaction on STEP Pretest



of this significant initial difference between treatment combinations on the criterion measure, a decision was made that the analysis of covariance would be the appropriate technique to be used for subsequent data analysis.

The problem of selecting the covariate (s) to be used was complicated by the nature of the criterion measure itself. The STEP Mathematics test consists of a variety of verbal mathematical problems and hence performance on the test might well be a function of reading achievement. To test out this possibility, a correlational analysis was made between the mathematics test and the two subtests of the Metropolitan Reading Test.[2] The correlations between the instruments are presented in Table 6. In interpreting the result, it should be noted that the correlation coefficients were based on class means and not on individual subject scores.[3] Inspection of the table reveals that the STEP pretest correlated higher with the Comprehension subtest of the Metropolitan than with the Word Knowledge subtest. The two subtests correlated quite closely with each other. In view of the latter relationship, a decision was made that both the Metropolitan Comprehension and Word Knowledge subtest scores along with the STEP pretest score would be used as covariates in any subsequent analysis.

---

[2] The mental ability test was not included in this analysis because of the early date at which it had been administered and hence was not considered as being highly successful. The Columbus Arithmetic Test was not included because of missing data.

Table 6.--Intercorrelations between Selected Pretest Measures

Variable	(1)	(2)	(3)
1. STEP Pretest	1.000	---	---
2. Metropolitan--Word Knowledge	.8856	1.000	---
3. Metropolitan--Comprehension	.9025	.9592	1.000

Analysis of Final Measures. Analysis of the final measures focused primarily upon the gains in mathematics achievement under the various treatment combinations of indirect and direct cues over the two year period of the study and the relationship between the exhibited gains and the measure of awareness as obtained at the end of the two year period.[4] Although the awareness measurement was obtained at the end of the two year period, analyses of the available data were made not only for the full two year period but also by examining data as obtained at the end of the first and second years respectively.

For purposes of data presentation, descriptive statistics of the individual classes are presented in Table 7. The table presents class means and standard deviations for the first, second, and two year gains in mathematics, means and standard deviations of awareness scores, and the gain-awareness correlations for the first, second, and full two years of the study. Data presented in the table are cited in presentations which follow regarding: (a) cues and gains and (b) gains and awareness.

---

[3] A further consideration is that the correlations were based upon the initial groups of students and not upon the final groups of students at the termination of the study as is the case with other analyses presented in this chapter.

[4] First year gain was STEP II--STEP I; second year gain was STEP IV--STEP III; the two year gain was STEP IV--STEP I.

Cues and Gains. Inspection of Table 7 reveals that the first year gain under each treatment combination tended to be similar with the exception of one classroom group under the C-D-T treatment. A similar observation can be made with regard to the gains during the second year of the study. For the full two years, the gains were also highly similar with the exception of one class under the C-D-T treatment and one class under the S-R-NT treatment, both of which were somewhat lower than for the other 14 classes. Based upon the 16 class means, the average gain for the two year period was approximately 12 raw score points.

To test the effects of the cueing factors or variables, separate analyses of variance of the true gain in mathematics for each time period were made. The summary values for analyses of variance are presented in Table 8. For convenience, only mean square values and F-ratios are presented. Inspection of this table reveals that no significant F-ratios were observed for the first year gain. The only significant F-ratio observed in the second year of the study was for the interaction between the Teacher and Informing factors, with this interaction being significant at the five percent level of confidence. A graphical representation of this interaction is presented in Figure 3.

For the full two years of the study, two significant interactions were observed. The interaction between Teacher and Method factors was significant at the one percent level while the interaction between Teacher and Informing factors was significant at the five percent level. Graphical representations of these two interactions are also present in Figure 3.

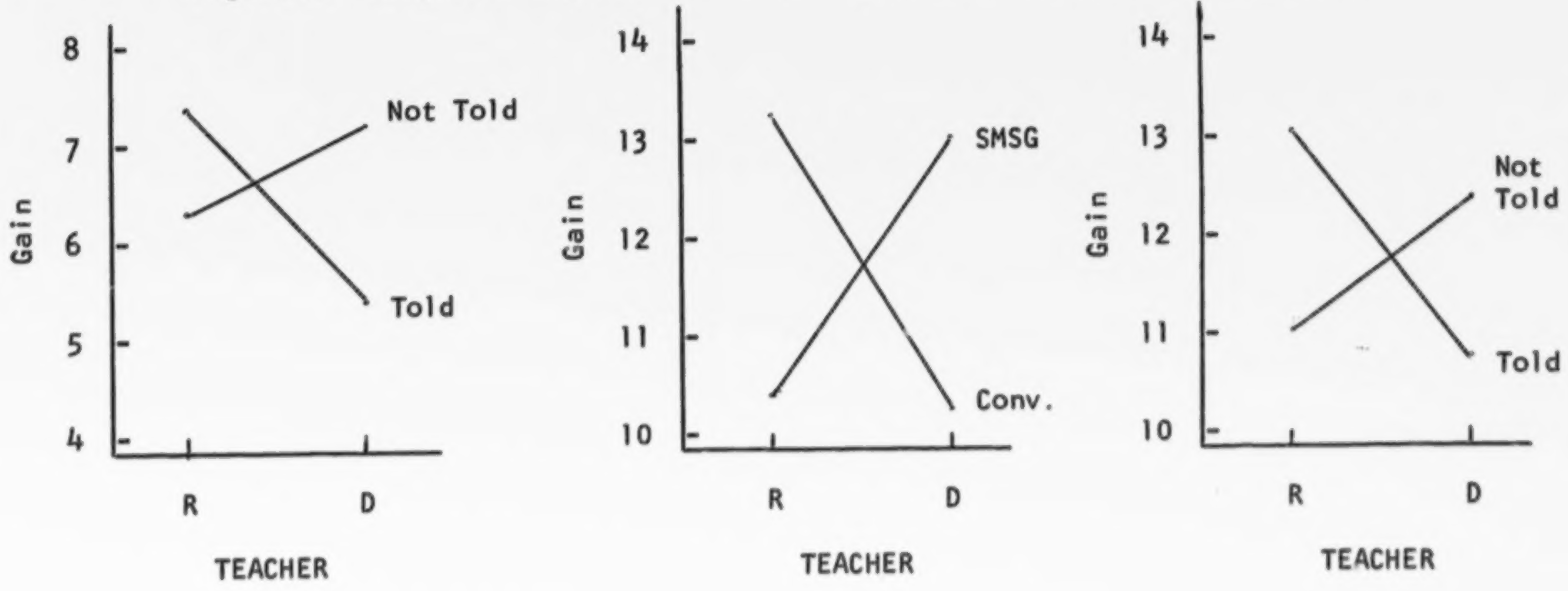
The general interpretation to be drawn from the F-tests just reported is that the variables or factors introduced as cues had no significant impact on subsequent student achievement. Under the common usage of the Hawthorne effect which states that such cues should cause differences in achievement, the main effect presented here does not support such a definition. It would seem reasonable to conclude, therefore, that the Hawthorne effect, commonly interpreted, operated in the experimental situation which was created expressly to test the commonly held hypothesis.

Table 8.--Summary Table for Analyses of Variance for True-Gain in Mathematics

Source	df	First Year I-II		Second Year III-IV		Full Two Years I-IV	
		MS	F	MS	F	MS	F
Teacher	1	0.200	.175	0.051	.206	0.432	.275
Method	1	0.166	.146	0.119	.482	0.170	.108
Informing	1	1.116	.979	1.121	4.544	0.058	.037
Teacher-Method	1	3.651	3.203	0.741	3.004	27.950	17.776**
Teacher-Informing	1	1.307	1.146	1.442	5.845*	11.197	7.121*
Method-Informing	1	1.317	1.155	0.001	.004	2.607	1.658
Teacher-Method-Informing	1	2.620	2.299	0.082	.332	1.120	.712
Error	7	1.140	--	0.247	--	1.572	--
Total	14	--	--	--	--	--	--

F(1,7;.05)= 5.59 F(1,7;.01)= 12.2

Figure 3.--Graphical Representation of Significant True-Gain Interactions



Second Year Teacher-Informing Interaction

Two Year Teacher-Method Interaction

Two Year Teacher-Informing Interaction

2

The significant interactions also contribute little or nothing to validate operationalizations of the Hawthorne effect. The significant Teacher-Method interaction noted for the full two years of the study was observed earlier in the pretest data analysis and therefore probably represents a continuation of the condition noted at that time. The significant Teacher-Informing interaction is harder to interpret because it runs counter to some commonly held ideas that informing students should perhaps be a stronger variable than changing the teacher. In this interaction, it can be noted that the Different Teacher was more influential than informing students that they were in an experiment. It should also be noted that a similar interaction did not appear in the first year gain analysis which might possibly be interpreted to mean that some changes among the participating teachers during the second year of the study introduced conditions which did not exist in the first year of the study. Whether chance or otherwise, they tended to carry over in a manner such that the full two year gain was effected thus causing the interaction to again appear in the latter analysis.

The results from the cues-gains indicate that the effort to establish an experimental classroom condition which would create differences in student achievement using specially introduced direct and indirect cues as independent variables was not successful. In terms of the objectives of the study, it may be concluded that any interpretation favoring purported presence of a Hawthorne effect under cueing conditions would not be supported by the data presented above.

Gains and Awareness. The Analysis of the relationship between awareness and gains was initiated by computing product-moment correlations between individual student gains in mathematics and scores on the awareness inventory for each class group. The individual class correlations between awareness and mathematics gains for the first year, second year, and the full two years were presented in Table 7. Inspection of the table reveals that the control group, as represented by the Conventional Mathematics--Regular Teacher--Not Told (C-R-NT) treatment, had the lowest awareness scores for the several treatments involved. The treatment combination of SMSG Mathematics--Different Teacher--Not Told (S-D-NT) had the highest average awareness scores.

Inspection of the correlations for the three different time periods reveals no consistent pattern of relationship existing between gains in achievement and awareness. Within the same treatment combination, individual classes exhibited both positive and negative correlations. This inconsistency in observed relationship is highlighted by the correlation of -.57 for one class within the C-D-T treatment and the correlation of +.57 for one class within the S-D-NT treatment, with both correlations being significant at the one percent level. For the second year of the study, only one correlation within the C-D-NT treatment was significant and that at the five percent level. Three correlations for the full two year data were significant at the five percent level. Again, no consistency was noted between the measure of awareness and the achievement gains under a particular treatment combination.

Using the two class measures under each treatment combination, mean scores for gain and awareness were calculated for each of the three time periods. These data are presented in Table 9. Product-moment correlations obtained between the mean awareness and gain measures for the first year, second year, and full two years were -.059, -.052, and -.136 respectively. The correlations were not significant at the five percent level of confidence.

The general interpretation of these results is that the observed relationships tend to be negative in direction and relatively low in magnitude and not significantly different from chance. These correlations indicate that students showing the greatest gains in achievement tended at the same time to be less aware of experimental participation, a somewhat direct contradiction of common Hawthorne effect usage.

It is recognized that the above results are based upon a questionable procedure since the awareness scores used were obtained subsequent to the time for measurement of gain, particularly for the first year of the study. The situation with regard to the second year of the study is more similar to the full two year situation since the awareness measurement was obtained at the end of the second year as were the two year gain scores. The full two year data does indicate a higher negative relationship than was observed for either the first or second year data.

The results presented in this section reveal no significant relationship between awareness of experimental participation and subsequent gains in mathematics achievement. From acceptance of the conceptualization of the Hawthorne effect as awareness of experimental conditions as presented in Chapter I, it would appear that the presence of such awareness is not related to subsequent achievement. If awareness is regarded as an independent variable and gain as a dependent variable, the

Table 9.--Treatment Means and Variances for Awareness and Gain Scores

Treatment	n	Awareness		First Year Gain		Second Year Gain		Two Year Gain	
		$\overline{X}$	s^2	$\overline{X}$	s^2	$\overline{X}$	s^2	$\overline{X}$	s^2
C-R-NT	2	33.40	.29	6.19	.76	5.27	.02	12.74	.79
C-R-T	2	51.81	.02	8.68	.13	5.17	.00	14.20	.17
C-D-NT	2	49.89	.00	6.84	.31	5.18	.00	11.93	.02
C-D-T	2	68.81	4.54	3.26	5.09	6.24	1.88	9.05	1.35
S-R-NT	2	55.09	1.40	6.83	1.45	4.51	.34	9.82	1.77
S-R-T	2	70.74	.02	6.54	.16	4.74	.01	11.95	1.23
S-D-NT	2	72.85	4.15	7.77	.00	5.27	.06	13.32	.66
S-D-T	2	70.41	2.20	7.96	.08	4.31	.00	13.09	.09

results here show that the presence of such awareness does not function as a causative factor of subsequent achievement. Accepting awareness as a definition of the Hawthorne effect, it would follow that the Hawthorne effect cannot be said to cause differences in achievement gains. Differences between treatment groups operating under different cueing conditions with regard to their awareness of experimental participation did exist. On this basis it can be said that the Hawthorne effect does appear provided that one accepts the conceptualization presented earlier in this report. The Hawthorne effect would not be present if it is regarded in the most commonly accepted usage of the term.

Duration of Study

One of the questions to be answered by the field research was whether or not extending the duration of a study could possibly be used as a procedure for controlling possible operation of factors normally associated with the Hawthorne effect. As noted in the methodology chapter, a decision was made to conduct the study over a two year period to see if achievement gains under the various treatment combinations would be significant at the end of the first year and then be reduced, remain constant, or increase over the full two years. Significant differences between treatment combinations at the end of the first year and nonsignificant differences at the end of the two year period might be used as support to the hypothesis that extending the duration of a study would reduce or minimize the operation of the Hawthorne effect.

To answer this question within the framework of the present study, repeated measurements using the initial STEP pretest for the criterion measure were made at the end of the first year and again at the end of the second year. The results were then analyzed separately for the first year and the full two year period.

It was noted earlier in the analysis of the pretest measures that there was a significant interaction between two of the treatment factors. In view of this interaction, the analysis of covariance was used to adjust the posttest means for initial differences in both mathematics and reading (28). Table 10 presents the individual class means for the first year raw scores on the criterion measure and the covariates while Table 11 presents the treatment means for the same variables. The mean scores presented for the criterion measures (STEP II and IV) are the unadjusted posttest means.

Table 10.--Unadjusted Class Means for First Year and Full Two Year Raw Scores on Criterion Tests

Treatment	Class	Covariates				Criterion				Final n*
		n	Metropolitan		STEP I	n*	STEP II	n*	STEP IV	
			W.K.	Comp.						
C-R-NT	1	24	29.00	24.29	17.63	24	20.54	20	28.60	17
	2	27	28.44	24.33	16.93	27	23.93	24	31.08	22
C-R-T	1	15	32.53	28.47	18.27	15	27.33	11	31.64	9
	2	19	35.63	29.53	18.21	19	29.05	18	34.56	17
C-D-NT	1	27	24.96	21.67	12.22	27	22.74	20	25.25	17
	2	26	24.00	21.38	15.00	26	23.96	15	27.80	10
C-D-T	1	30	25.13	24.47	16.57	30	22.43	23	25.48	19
	2	23	19.17	18.44	11.26	23	18.13	18	21.94	17
S-R-NT	1	32	23.28	17.18	12.13	32	21.44	23	22.56	20
	2	30	22.48	8.80	14.53	30	18.30	24	20.46	22
S-R-T	1	26	26.04	22.50	12.27	26	20.19	20	28.30	15
	2	27	25.19	21.26	15.37	27	19.70	17	24.06	17
S-D-NT	1	25	32.92	28.96	18.64	25	26.60	20	30.40	16
	2	26	31.56	28.64	16.88	26	25.36	20	30.95	18
S-D-T	1	27	33.78	31.04	18.41	27	26.44	19	31.90	16
	2	27	32.15	29.82	20.07	27	29.15	26	33.50	26

* n=number of subjects used for analyses of covariance reported in subsequent tables except for final n which was used for awareness-gains analyses.

Table 11.--Unadjusted Treatment Means for First Year and Two Year Raw Scores on Criterion Test

Treatment	n	Covariates			Criterion	
		Metropolitan		STEP I	STEP II	STEP IV
		W.K.	Comp.			
C-R-NT	2	28.72	24.31	17.27	22.23	29.84
C-R-T	2	34.08	28.99	18.24	28.19	33.10
C-D-NT	2	24.48	21.53	13.61	23.35	26.53
C-D-T	2	22.15	21.45	13.91	20.28	23.71
S-R-NT	2	22.88	17.99	13.33	19.87	21.51
S-R-T	2	25.61	21.88	13.82	19.95	26.18
S-D-NT	2	32.24	28.8	17.76	25.98	30.68
S-D-T	2	33.96	30.43	19.24	27.79	32.70

Summary tables for the analyses of covariance for the first year and the two year mathematics raw posttest scores are presented in Table 12. Inspection of the table reveals that there were no significant differences at the end of the first year either with regard to main effects or interactions. The general interpretation of this result is that the treatment factors, taken singly or in combination, did not produce any significant impact upon the posttest measures.

A similar analysis was undertaken with regard to the full two year mathematics raw scores. Table 10 presents the unadjusted classroom mean posttest scores for the criterion (STEP IV) at the end of two years with the treatment means presented in Table 11. The summary table for the analysis of variance for the full two year mathematics raw scores presented in Table 12 indicates a condition similar to first year scores in that there were no significant main effects or interactions.

These results indicate that a suggested hypothesis of significant differences in achievement during the first year and then leveling off during the second year is not supported. The general finding is that extending the study for a two year period did not seem to make any notable contribution to justification for an extended duration of a study as a technique for controlling the presence of a possible Hawthorne effect.

Table 12.--Summary Tables for Analyses of Covariance for First Year and Two Year Raw Scores on Criterion Tests

Source	d.f.	First Year				Two Year			
		S.S.	M.S.	F	Decision	S.S.	M.S.	F	Decision
Teacher	1	7.873	7.873	2.989	NS	0.79	0.79	0.23	NS
Method	1	3.739	3.739	1.420	NS	3.47	3.47	1.02	NS
Cueing	1	0.644	0.644	0.245	NS	0.00	0.00	0.00	NS
Teacher-Method	1	0.000	0.000	0.000	NS	1.62	1.62	0.47	NS
Teacher-Informing	1	0.707	0.707	0.269	NS	2.24	2.24	0.66	NS
Method-Informing	1	1.056	1.056	0.419	NS	5.56	5.56	1.64	NS
Teacher-Method-Informing	1	3.429	3.429	1.303	NS	0.96	0.96	0.28	NS
Regression	3	13.508	4.503	--	--	12.50	4.17	1.23	NS
Adjusted Error	5	12.166	2.633	--	--	12.50	4.17	1.23	NS
ANOVA Error	8	26.674	3.334	--	--	29.49	3.69	--	--

F(1,5;.05)= 6.61 F(1,5;.01)= 16.26

Teacher Interview Results

As noted in Chapter II, one project evaluative technique employed was a post-experimental inquiry or interview of the participating teachers. The purposes of these interviews were (1) to secure any evidence about the possible existence of the Hawthorne effect which might not be assessed through the measurement of achievement gains and subject awareness and (2) to secure possible information which might relate to the success or failure of the treatment conditions. The results of these interviews follow.

The general nature of the teachers' comments seemed to be quite closely related to the traditional interpretation of the Hawthorne effect phenomenon. For example, a teacher in one school reported that as the students were carrying their new textbook up and down halls (when it was rather unnecessary to do so), she noticed that they seemed to be doing it because the students were developing an espirit-de-corps in the experimental classrooms as a consequence of studying "mathematics" out of a new book while other classrooms were studying the "same old thing."

During the series of interviews the teachers reported that, although the teaching of mathematics in any classroom could be a formal activity, the suggestions from the Associate in Mathematics and from the teacher's edition of the SMSG textbook provided adequate suggestions and gave the teachers of the new method sufficient material to develop a feeling that they were working closer to and more continuously with the way a child chose to learn a new method or concept. The teachers also felt that they could guide the learning experiences of the children better under the new method but that it was costing them a little free time they had formerly enjoyed during more frequent periods of individual study.

There were some teachers who believed that student acknowledgment of novelty, or attention, or variation in his environment might provide a positive motivation to increase his criterion achievement scores. The possibility existed, however, that it may have also provided no increase or may have even provided a negative motivational factor. The latter situation might be indicated from the report of two teachers in one school who reported that in the middle of the fourth grade and again during the fifth grade that their students indicated a rather strong resentment at "being required to study this hard material." Whether or not such a reaction did indeed provide a negative motivational factor is up to the reader to ascertain

Assuming that it would, one might have difficulty in predicting its influence on the criterion variable of achievement.

One major finding of the teacher interviews was that the use of different teachers did create some degree of novelty at the beginning of the fourth grade. According to teacher reports, this situation appeared to last about three or four weeks and was of sufficient impact to be reported in the experimental log maintained by the Associate in Mathematics. The increase in novelty seemed to be connected more with the students' testing of the discipline under the different teacher than with the novelty of having a different teacher in the room. In one school, the novelty effect surrounding the different teacher seemed to last considerably longer than in the other schools. The daily time selected for changeover of the teachers in this school, however, coincided with a recess break as well as a milk period and hence was complicated by a scheduling variable unique to this one school.

The teachers in the SMSG classes reported that parents seemed to react more to the introduction of the new mathematics than did the students. This finding agrees with a report by Duncan (23) in which a parental in-service program provided a significant increase in student growth. The way in which most parents expressed their interest, positive or negative, was in a manner that indicated they were more concerned personally with not being able to help the students with their homework than in the change of method per se.

In both of the SMSG--Not Told schools, the teachers thought that telling the parents or children they were in an experimental classroom might have eased them over the initial burdens created by the introduction of SMSG mathematics. At the same time, the teachers felt that this would not have solved the parent-student problem without some other aid. In one school, the teachers reported that the parents were glad that their children were in a new mathematics program since they were headed for a junior high school that was teaching the same program. Students currently in the junior high school were reported as having trouble because of their different elementary school backgrounds. These same teachers reported that even the children indicated that they liked the new mathematics since it was now possible for them to help their older brothers and sisters with their junior high school homework.

The teachers in one school reported that the children as fourth graders had developed a sort of group spirit about being able to have arithmetic from a different teacher. The students were described as building this feeling into a type of competition between classes without the teacher's assistance. During the fifth grade, these same students expressed a resentment about the intermixing of classes since their group lost its identity for arithmetic competition.

None of the teachers reported that they had even been directly involved in an experimental study in the past. Several did report that other teachers in their schools had been in such studies but not of such a strong comparative nature. The teachers expressed the feeling that the adaptation to the conditions implied by the experimental treatments imposed unusual hardships. This was particularly true where there was a need to be familiar with the new SMSG materials or where the teaching of different classes required a close adherence to a time schedule.

In several of the schools, teachers reported a feeling that the two years of the experiment should have continued with the same classes of students intact in order to benefit from group similarities and a continuity of effort for the similar groups. Some other teachers felt that the annual shifting of students from class to class should be on as wide a basis as possible in order to avoid a formation of school cliques and low motivation groups.

The teachers at one school reported that the students had quite evidently separated their answers to the awareness scale in a manner which showed definitely that they were aware that it was the class and not the school that was in the experiment. The students also showed more than a passing interest by noting that many of the statements seemed to be repeats of earlier ones. Some students were described as being reluctant or refusing to mark several of the personality items included on the awareness measure.

The purpose of this chapter was to present the results of the field research phase of the total project as they pertain to the objectives and questions concerned with the role of direct and indirect cues in producing the Hawthorne effect, the relationship between achievement gains and awareness of participation in an experiment, and the duration of a study as possible control procedures for the Hawthorne effect. A discussion of these results is presented in Chapter V. Chapter IV following presents results with regard to the literature analysis phase of the project.

Chapter IV

Literature Analysis Results

Introduction

As was noted in Chapter I outlining the objectives of the study and again in the chapter on methodology and procedures, one projected source for securing additional information about the phenomenon under study was a systematic analysis of published research reports dealing with comparisons of new versus old methods of instruction and/or curriculum evaluation. The purpose of this chapter is to present the results from the literature analysis undertaken as part of the total project.

The questions to which this section of the total report relates were questions 1, 2, and 4 under general question A. It was anticipated that the information obtained might assist in answering the questions under general questions B and C. Question 4 was further broken down into two separate questions for purposes of the literature analysis. It was not possible to collect sufficient data in the form of abstracted studies which would materially assist in arriving at any answer for question 2. Therefore, the report does not contain an answer to this question. There were, then, three separate questions involved in this part of the total project. The chapter has been divided into three principal parts or sections to present the results of the literature analysis phase.

The first part attempts to answer question 1 and presents definitions of the Hawthorne effect as set forth by various investigators along with their suggestions for control. It also presents evidence relating actual control procedures used to experimental results in order to determine the effectiveness of such procedures. The second part relates to question 4 and presents an analysis of the relationship between experimental results and the duration of time that a study was conducted. The third part also relates to question 4 and presents an analysis of relationships between changes introduced within an experimental situation and the results of the studies as presented by the investigators. Each part presents the purpose, procedure, and results relating to it. A general discussion of the results is presented in Chapter V.

Control Procedures for the Hawthorne Effect

Definitions and Suggested Controls

Purpose. While many investigators have used the Hawthorne effect concept to explain unexpected or even undesired results, the concept still eludes definition. Attempts at definition by investigators have produced varied meanings while in some cases no attempt has even been made. Even though there is some confusion about what the Hawthorne effect _is_, there have been numerous attempts to control for it nevertheless. The purpose of this section is to present the results of an attempt to relate the definition of the Hawthorne effect used by an investigator to the suggested techniques to be employed for control of it. If studies employing the same definition propose different means of control or, if studies which define the effect differently nonetheless use the same means of control, such inconsistencies should be noted in any attempt to understand the concept.

Procedures. The studies used in this analysis were drawn from the research abstract file described in Chapter II. From these abstracts, only those studies which specifically mentioned the Hawthorne effect and discussed control for its influence were chosen for inclusion in this analysis. There were numerous other study abstracts which mentioned efforts to control for the "novelty" effect, the "placebo" effect, or similar terms which are often used as synonyms for the Hawthorne effect. However, because of the confusion which could arise as a result of vague definitions it was decided not to include these abstracts. The Hawthorne effect term _per se_ had to be used for the study to be included.

The first step was to identify the definitions of the Hawthorne effect used by those investigators who specifically mentioned an attempt to control for it. Of the twenty-five abstracts which met the criterion noted above, only nine contained any explicit attempt at definition. As might be expected, the definition of the Hawthorne effect set forth by these investigators usually determined the variable they considered necessary to control. For instance, if an investigator defined the Hawthorne effect as a motivational effect in nature, he usually attempted to control motivation in his experiment. Sixteen studies did not include an explicit definition of the Hawthorne effect but did note suggested controls. Thus the variables which the investigators controlled were inferred for the purposes of this analysis.

Results. Of the twenty-five studies, twenty-one could be grouped into four categories corresponding to the general variable said to be controlled. These variables were attention, motivation, novelty, and awareness of an experimental situation. The other four studies eluded grouping since, for the most part, although the investigators specifically mentioned an attempt to control for the Hawthorne effect, in no case did they indicate what variable they believed should be controlled.

Attention. The variable which most investigators considered important to control was that of attention. The focus of attention however was inconsistent. Some believed that it was special attention paid to the teachers of experimental groups which contaminated the results, while others believed it was the attention paid to the students. Still others believed it was necessary to control attention paid to both teachers and students.

Oberholtzer (54), Brickell (8), and Wolf, et al. (83), considered teachers to be proper recipients of control. Oberholtzer stated that similar treatment should be paid to teachers of both the control and experimental groups. He believed this could be done by equalizing the number of in-service meetings planned, the amount of assistance given, and the number of curriculum bulletins issued. Brickell stated that teachers of both the experimental and control group should receive the same amounts of attention, encouragement, and recognition given by people from outside the classroom. He noted that,

> When these are supplied equally, overproduction should occur in both the comparison and experimental groups. Then the differences in results, if any, can be attributed with more confidence to differences in the program--rather than simply to the fact that something was changed. (8, p. 35).

In a project proposal by Wolf, et al., a summer workshop was planned for teachers of experimental groups. In order to equate amount of attention paid, a similar workshop was designed for control teachers. It was further planned that both groups of teachers would receive the same amount of materials and instruction during the year-long experiment and would attend the same number of conferences.

Rimoldi, et al. (64), Forehand (27), and Entwisle (25) considered students to be the necessary recipients of the control of the attention variable. Rimoldi administered the Rotter Incomplete Sentences Test and the Rosenzweig Picture Frustration Test to control subjects in order to equate their treatment with that of the experimental subjects who were given problem solving materials. Forehand, in an article reviewing research on the teaching of English, stated merely that the experimenters should "...make the control course as 'special' as the experimental course, or the experimental course as 'routine' as the control...." (27, p. 54). He did not elaborate on how one should go about doing this. Entwisle stated that an active attempt to control the attention variable was unnecessary so long as a method of measuring the amount of effect this variable has on the results is available. She suggested that,

> By withholding some of the experimental material entirely from both groups, spurious gain of the control group on the experimental material can be assessed, for without communication the control group should gain the same amount (within sampling fluctuations) on the materials shown to the experimental group as well as on equated material of the same type shown to no one. (25, p. 478).

In a monograph titled Psychological Research in Education (62) published by the National Academy of Sciences, both the students and the teachers are considered the correct recipients of control. It is stated that equal amounts of attention should be given to all treatment groups, but there is no further elaboration. Veatch (80), in reporting a study done by Cyrog,[1] did not mention who the recipients of control are or should be, but did state that attention can be controlled by equalizing the amount of time in class spent by both groups, the class size, the consultant assistance paid, and the administrative emphasis.

Motivation. It was possible to identify another group of investigators all of whom stated that it was motivation which needed to be controlled. Within this group, it was usually the case that the Hawthorne effect itself was defined specifically as being a motivational effect. Porter (61), Bowers and Soar (7), Rosenbloom (67), Hershberger and Terry (31) considered the students to be the appropriate recipients of this control. Porter defined the Hawthorne effect as "...the motivational results of merely engaging in an experiment irrespective of the differential treatments carried out...." (61, p. 37). He stated that in a study concerned with methods of teaching spelling, the Hawthorne effect was the direct result of social intercourse between students. Because it was believed that direct measures of controlling this social intercourse would act to stimulate the Hawthorne effect, they were not undertaken. Instead, anecdotal records of the classroom behavior were kept. In the study by Bowers and Soar on human relations in the classroom, the teachers were the students at a summer workshop. In order to control for differential motivational effects, a summer college course of the teacher's choosing was offered free of charge to the control teachers. Besides this an attempt was made to communicate equally with both experimental and control teachers about the nature of the experiment. Rosenbloom, in a discussion on methods of teaching mathematics, advanced the opinion that the best way to equalize motivational effects of an experiment would be to make the control method appear experimental. According to Rosenbloom, the best way to do this would be for the conventional texts to be lithographed and stamped "Experimental Edition." This procedure was not possible in the experiment cited and instead the conventional text was supplemented by a lithographed pamphlet. In the study by Hershberger and Terry, the regular teachers administered the experimental materials. They were instructed not to inform their students that they were in an experiment and not to urge the students to excel. The unusual aspect about this study was that the Hawthorne effect was defined as a novelty effect although the variable controlled for was motivation. The increase in motivation that it was necessary to control was apparently not conceived by the investigators as stemming directly from the novelty of the situation but rather from the knowledge of being in an experimental situation and from the teacher's attitude. In this sense, the teachers as well as the pupils would be appropriate recipients of the control.

---

[1] It is Veatch, not Cyrog, who mentions the control of the Hawthorne effect *per se*.

Other investigators who considered motivation to be the variable to be controlled did not conceive of the students as the appropriate recipients of this control. Payette (56), in a study of methods of teaching mathematics, stated the opinion that it is the teachers, not the students, who should be the recipients of the control. In the experiment which he reported, a control-control group was used. This group did not know they were to participate in an experiment until just before the administration of the final tests. Pella, et al. (58), in a study of the teaching of physics, also used the control-control group technique but these investigators considered the students as well as the teachers to be the recipients of control. Lambert, et al. (40), also considered both teachers and pupils as the correct recipients of control. Motivation here was controlled in two ways. First, only those teachers who were strong advocates of whichever method of teaching they were using were utilized. Second, both the experimental and control groups were stimulated by special visitors, new teaching aids, and so on. Thorndike (79) also considered motivation as the variable to be controlled but does not mention whom he considered the correct recipients of the control. His reference to the control of the Hawthorne effect is not in the context of an experiment. It may be for this reason that he was vague. He stated that in order to control for the motivational effect, the control treatment as well as the experimental should be given the "flavor of an experiment" but did not elaborate on this suggestion.

Novelty. Three investigators reported "novelty" to be the variable which has to be controlled. Borg (6), who believed that both teachers and pupils should be the recipients of this control, stated

> The influence of Hawthorne effect can be expected to decrease in the school situation as the novelty of the new method wears off, and therefore, studies extending over a period of two or three years can be relied upon somewhat more.... (6, p. 339).

Pella (57) agreed that it was novelty that should be controlled but disagreed as to method. He did not think the length of the experiment to be the important factor. Rather, he believed that it was most important that the teachers become proficient in the use of the experimental methods and that the students become accustomed to their use. Higgins and Rusch (33), in a study of programmed arithmetic texts, used special materials in the control classes as well and impressed upon all the students, experimental and control, that they were taking part in an important experiment.

Awareness. Two investigators considered awareness of an experimental situation to be the variable to be controlled. Sjogren (74) conducted a study concerned with the presentation of correspondence courses. In order to control for awareness of the experiment, he sent similar materials to people in the same geographical areas to minimize the possibility of discovery that there were different materials. Downing (21), also stated that he considers it necessary to control the awareness of the experimental situation. His method of control, however, appeared more as one to control for motivation.

> We have attempted to make it clear to the control group (of teachers) that we are just as interested in its work as we are in the work of the experimental group. We match our visiting schedules on the two groups. We try to make control group meetings as

stimulating in content as those of the experimental group.[2]

Unclassified. Four studies mentioned the Hawthorne effect and offered a method of control which could not be placed in the four operational categories already discussed.

In one such experiment by Crosby, et al. (20), each teacher taught both the experimental method and the conventional method. For this reason, it was believed that the variable important to control was that of preference for one technique over another. The teacher's verbal statement of no bias was accepted for this purpose.

Robinson (65), Cooper (17), and Lang and Tohtz (43), presented papers in which, while stating a method to be used to control for the Hawthorne effect, they never discussed what specific variable they considered to be the necessary one to control. Robinson merely stated that, "...careful planning and thinking through of the total design and possible outcomes might help investigators prevent at least some of the Hawthorne effect." (65, p. 172). Lang and Tohtz in a report of a programmed learning experiment used a control-control group and Cooper, in his experiment of English instruction for bi-lingual children, used a "Non-Experimental Control" group which was essentially the same as a control-control group. Neither study stated nor indicated what variable these groups were designed to control.

Summary. The twenty-five studies found which mentioned controlling for the Hawthorne effect were grouped into five categories--attention, motivation, novelty, awareness of experimental situation, and miscellaneous--according to the definition of the Hawthorne effect which was offered by the investigator or which was inferred as a result of the variable controlled.

Eight studies belonged in the attention category. Of these, five considered it necessary to equalize the treatment given to the experimental and control teachers or pupils; one stated that control conditions should be made to appear experimental through the use of additional materials; and two studies did not hypothesize any specific method of control.

---

[2] Personal communication to Principal Investigator.

Eight studies considered motivation the important variable to control. Six methods of controlling motivation were used. Of these, two studies suggested that similar treatment be given to experimental and control groups. This was also the most popular method used by the "attention" group. One study hypothesized that the control condition be made to appear experimental; one study stated that inter-student communication be controlled; two studies used control-control groups, one study suggested experimental conditions be made to appear ordinary; and one study did not suggest any method.

Three studies considered novelty the important variable to be controlled. Of these, one suggested it was the duration of the experiment that should be controlled; another considered the proficiency of the teacher in the new method to be the important factor. The third suggested that control conditions be made to appear experimental.

There were two investigators who believed the awareness of the experimental situation to be the variable to be controlled. One suggested awareness be controlled by minimizing inter-student communication. The other considered it important to give similar treatment to both the experimental and control groups.

Of the four studies which couldn't be placed in the above categories, two used a control-control group to control for the Hawthorne effect; the third suggested that teacher preference be controlled; and the fourth did not offer a specific method of control.

Of the twenty-five studies, ten considered the students to be the correct recipients of control, six considered the teachers, and six considered both the teachers and the students. In the other three studies, the recipients of control were not indicated.

It can be seen from the above information that not only was there great inconsistency of definition but also wide variability of suggested methods for controlling the Hawthorne effect phenomenon. Investigators who defined the effect similarly often use different means of control while the same means of control were sometimes used by investigators who defined the effect differently.

Control Procedures as Related to Experimental Results

Purpose. The previous section was devoted to definitions of the Hawthorne effect as set forth by various investigating and suggested variables that should be controlled to reduce its potential effects on the experimental situation. Further inquiry about the role of the phenomenon was made by examining the way investigators did in fact try to control for possible operation of the phenomenon in actual experimental situations and the relationship of such controls to the subsequent results. The purpose of this section is to present the results of an analysis of selected literature to determine the relationship between actual control procedures employed and experimental results.

Procedure. The studies utilized for this analysis were drawn from the research abstract file with only those studies specifically using the term Hawthorne effect and employing a control procedure for it being included. Sixteen studies were within the explicit criteria used.

The type of control used in each study was determined initially from statements made in the original reports. These controls were tabulated and then grouped in two major classifications with specific variations being listed under each classification.

For each study used, information gathered for this part of the investigation included age level, number of subjects, duration of the study, hypotheses tested, criterion measures used, statistical measures applied, and the author's conclusions of significance. From this information, a general statement of results was derived for each study and placed into one of three categories: E>C, E=C, or E<C. Where several hypotheses were tested, these results were also recorded in a similar manner.

Results. The major classifications of control methods subdivided by the specific variations and the corresponding results are summarized in Table 13.

Inspection of the table reveals that there is no definitive relationship existing between the type of control used and the general statement of results. In view of the absence of an overall relationship, the effectiveness of the specific variations of the type of control was studied by examining each investigation on an individual basis.

Use of More Than One Type of Control Group. Four studies employed more than one type of control group. Pella, et al. (58),

Table 13.--Control Procedures as Related to Results

Control Procedures		Results		
A. MORE THAN ONE CONTROL GROUP		E>C	E=C	E<C
1. Control-control group established after experimental period has lapsed. CC has no knowledge of the experiment.	C>CC	0	0	1
2. Control-control group established with E and C groups but not told of the experiment.	C>CC	1	1	0
3. Control-control established with E and C but C receives pseudo-treatment.	C=CC	0	1	0
B. EQUALIZATION MEASURES		E>C	E=C	E<C
1. Experimental and Control groups not told of participation in an experiment.		2	1	1
2. Experimental and Control groups both told they were the E group.		0	1	0
3. Activities employed to equalize motivation.		2	1	1
4. Duration of experiment assumed to eliminate Hawthorne effect. (novelty definition)		0	1	0
5. Experimental school--assume students perceive experimentation as ordinary environment.		0	1	0
6. Control group receives pseudo-treatment.		0	1	0

used a control-control[3] group established after the experimental period was over. This group presumably had no opportunity to be influenced by experimental participation. The experimental group was taught by the Harvey White Physics film series and the control groups were taught as usual (i.e., without Harvey White Films). Achievement was measured by standardized subject matter tests of physics. Duration of the experiment was one 90-day semester and the sample consisted of 63 Wisconsin High Schools in cities under 100,000 population. Results showed that "...the control groups out-performed both the Film and control-control groups by 2.5 and 2.05 respectively. These differences were significant at the .05 level when tested by analysis of covariance." The authors commented on this outcome.

> Because the C.C. groups neither saw the films nor knew that they were in the experiment (until five days before taking the semester-end test), it might be assumed that the superior performance of the control groups was the result of the "Hawthorne effect" (increased enthusiasm and effort due to groups knowing they are the control group in an experimental situation). (58, p. 11).

The CC method was also used by Cooper (17) in accounting for the Hawthorne effect. In this study, the CC group was established with the experimental and control groups but were not told of the experiment. The purpose of the investigation was to determine the effects of different amounts of first grade oral English instruction upon later reading progress with Chamorro-speaking children. The experimental groups were Chamorro-speaking children exposed to one year of conversational English in grade one rather than the conventional basic reader in English (E_1), Chamorro-speaking children exposed to one-half year of conversational English in grade one instead of the basic reader (E_2), pupils instructed in English reading with the standard basic reader approach from first grade on (C_1), and pupils in five additional classes from other schools not associated with the experiment and not informed (CC). Except for CC group, the subjects were first grade pupils in four schools each containing two experimental classes and one control class. Measures were taken of speaking ability at the end of grade one and reading comprehension at the end of the fourth year. Results showed that in speaking ability $E_1 > E_2 > C_1 > CC$ and in reading comprehension generally $E = C$. Again differences in achievement scores between control groups appeared. The author commented,

---

[3] The symbol CC is used hereafter to refer to the control-control group concept.

> Two sets of control classes were studied; one in each of four experimental schools, and five in schools unrelated to the study. This last was to see if "Hawthorne" type effects developed. They did not. (17, p. 125).

An evaluation of the School Mathematics Study Group curriculum was conducted by Payette (56) using a CC group also established with the experimental and control groups yet not told of the experiment. However, in this study the attention was given to the teachers attitudes and preferences and the experimental control groups were established on this basis. Teachers of the CC group were not informed about the experiment. Achievement in mathematics was measured for groups exposed to SMSG materials from teachers inexperienced in teaching SMSG (EA), SMSG materials from teachers experienced in teaching SMSG (EB), experienced conventional teachers (CA), experienced conventional teachers not informed of their participation in the experiment (CC). The subjects were students in mathematics classes in grades 7-12 in seventy-five schools in the United States. In this instance, results reported were that CA=CC and E>C. The author commented as below.

> Comparisons of achievement on conventional and SMSG tests for students of CA and CC teachers indicate unequivocally that there is no advantage in favor of students of CA teachers, those teachers who knew they were in an experiment. (56, p. 11).

Lang and Tohtz (43) used a CC group in another manner in their study. The subject was the effectiveness of out-of-class-programmed instruction in English in enhancing achievement in recognizing correct English usage and in writing essays. Here, the control group was given a pseudo-treatment to match treatment influences with the experimental group while the CC represented conventional classroom routine. Results in this study showed in general that E=C=CC.

In these latter two studies, where results showed control equal to CC in contrast to the previous examples where control was greater than CC, the common element of special considerations of teacher attitudes and backgrounds raises an intriguing question about the influence of this factor when a CC group is used in any manner for special control.

Equalizing the Influences of the Experimental Situation Between All Groups. The first variation under this classification is where both experimental and control groups are not told of participation in an experiment. Four studies used this method for control. One study conducted by Sjogren (74) showed results where $E_1=E_2=C$. The experimental group used a programmed text in ninth grade algebra and the control group received the usual materials for a correspondence course. The total N was 186 with over half the subjects having overseas addresses. Another study using this method of control was conducted by Hershberger and Terry (31). This study investigated the effectiveness of varying degrees of typographical cueing in both programmed and conventional texts. The variables were no typographical cues, simple cues, complex cues, and complex cues without directions. Each variable was tested in a programmed text using experimental groups and in a conventional text by use of control groups. The criterion measure was gain scores on pretest-posttest schedule of objective tests covering the history of Texas. Results show that $E>C$ and that E_2 and $C_2>$ all others.

In a study by Ryan (71), one of the several controls used for the Hawthorne effect was not informing either the experimental or control group of the experiment. In addition, the instructors were selected on factors that tended to equate teacher motivation and minimize bias. Each instructor taught both experimental and control groups. Also advantageous was the fact that variation in instructional approach was the common environment at the participating school. Comparisons were made between the achievement of students given a choice of three different instructional methods (E group) and those given no choice of method (C group) in a course in Educational Psychology at Oregon State University. Tests of subject matter knowledge presented in class and application of principles were given and results indicated that $E>C$ in both measures. Keating (38) also had this variation of control built into a study on the effectiveness of language laboratories by selecting subjects who were all in on-going programs and had no knowledge that the test results would be used for comparison purposes. The subjects were more than five thousand French secondary students at four levels in twenty-one school districts of the Metropolitan School Study Council. Measures were taken on reading comprehension, listening comprehension, and speech production. The experimental group was in programs using a language laboratory and the control group in programs where no language laboratory was used. Results recorded show $C>E$ except for speech production in first year French where $E>C$.

The opposite approach was used by Higgins and Rusch (33) in evaluating the use of programmed texts and workbooks in remedial arithmetic classes of fifth and sixth graders. The authors explained

> Hawthorne Effect Controlled: All the children were told they were to take part in an important experiment in which two methods of learning would be compared. Arithmetic workbooks had never previously been used in these classes. Every child in the fifth and sixth grades was told he was an experimental subject. The principal visited each class and impressed the youngsters with the importance of the experiment. Other special visitors repeated the statements made by the principal. Since all students were reinforced by authority figures and since workbooks had never been used before, it was hoped that the novelty effect of the programmed text would be counter-balanced. (33, p. 34).

The sample of 78 subjects were matched on 7 factors including IQ and Reading Comprehension in addition to subtests of various arithmetic skills. Criterion measures were four standardized tests of achievement in arithmetic. Their results showed that E=C.

Four studies were placed in another category of variation where the common element was the application of activities to equalize the motivation of those persons involved in the experiments. However, each study employed different specific activities to accomplish this goal based on the nature of the design of each study and the author's anticipation of situations where obvious inequalities in motivation would occur between participating members. In all four studies, these activities were directed in the large part to the teaching staff involved and usually took the form of teachers' workshops, special interviews, and periodic stimulation by the investigating staff.

One study by Downing and Hallwell (22) was still in progress at the time of the analysis with the results reported being those recorded at the end of the second year of the project. These tentative results showed E>C in rate of learning, vocabulary, accuracy, comprehension, and spelling. The study compared Initial Teaching Alphabet (ita) for use in beginning reading to the conventional method using the traditional alphabet. The study was conducted by the Reading Research Unit of the University of London Institute of Education using a sample of 220 pupils in British Infants Schools.

Bowers and Soar (7) analyzed teacher attitudes and personality after a summer workshop consisting of intensive laboratory training in human relations. Changes were measured by observation of classroom behavior of teacher and pupils and by standardized personality and attitude measures. The experimental group consisted of elementary teachers who participated in the workshop while the control group was elementary teachers who had the option of taking a college course of their own choosing at no expense. The results of this study show that E>C in attitude change toward pupils. In classroom behavior the tendency was E>C with some variation.

Lambert (40) also employed this procedure as noted by the statement:

> Care was taken to introduce bias in the selection of staff for this study: only strong advocates of the team and self-contained organization were retained or employed.
> Each group was stimulated by university consultants and the introduction of new teaching aids, that is, overhead projectors, teaching machines and tape recorders. All groups were stimulated to counteract possible Hawthorne Effects. (40, pp. 33-34).

The experimental variable in this case was a team-teaching organization at the elementary school level compared to a modified self-contained organization. The modification was that the students in the latter group were also exposed to specialists in the fields of physical education art, and music. The sample consisted of one class at each grade, first through sixth, with the total number of subjects being 350. The setting was two elementary schools in Madison, Wisconsin with one school containing both an experimental and control group and the other control group only. The duration of the study was two school years. Group differences investigated were group interaction where E=C; personality adjustment where E=C (C>E at first year measures); and achievement where C>E.

Crosby, et al. (20), had as the independent variable the use of a new teaching method of individualized instruction. The new method structure required pupil participation in individual goal setting, planning their work within the framework of the syllabus and their interest and abilities, requesting tests when ready, and participating in the evaluation of their work. The control conditions consisted of a conventional teaching method of the teacher controlling goal setting, work assignments, and testing and evaluation. The dependent variables consisted of pupil achievement in algebra and end-of-the-year attitude toward mathematics. The sample group came from 36 ninth grade elementary algebra classes in nine junior high schools in the New York metropolitan area. The total sample was 904. The duration of the experiment was nine months, or one school year. Specially constructed instruments were used to measure achievement and attitude toward mathematics. Results reported in achievement were E=C and in attitude E=C with mean scores lower for both groups at the end of the year. Special considerations made for control purposes were a persuasively constructed first interview with each teacher which emphasized that preference for one method over another would adversely affect the experiment and a summer institute to train the experimental teachers in the new method to minimize preference for the conventional familiar method of teaching.

An effect from novel stimuli has often been noted in conjunction with or mentioned as a component of the Hawthorne effect. The logical assumption that the novelty effect would decline in intensity as time passes, if valid, would have an important influence in determining an optimum duration of studies. Porter (61) pursued this question in a study and found that

> Over the entire period of the experiment a drop in achievement was observed, but it was no greater on the teaching machine lessons than on teacher-taught lessons, and the trend was toward a smaller loss on the machine lessons. The drop in achievement appears to have been a result of general classroom operations and administration of the programs rather than an inherent characteristic of the teaching machine instruction. (61, p. 108).

The study was instituted to compare the achievement results of elementary level pupils in spelling using teaching machine instruction as the independent variable contrasted to teacher-taught lessons. The duration was one school year. Significant differences were observed the first six months, however, the differences were not significant during the second six months and were not significant over the full year.

Another study investigating the influence of novelty was conducted by Popham (60). This research was carried out in a college demonstration school where experimentation and observers are the ordinary environment. The author felt that the very uniqueness of teaching machines may heighten the student motivation and thus engender superior learning. It was also felt that since most experiments with teaching machines were of such short duration that perhaps if they were extended the influence of the novelty effect might lose its value. Two groups of sixth grade students from a college demonstration school were taught some fundamentals of algebra during the fall semester of 1961-62. One group was taught by use of a teaching machine and one group by means of regular classroom teaching. During the second semester of the 1961-62 school year, the members of both groups were randomly assigned in equal proportions to four teaching machine work groups and all were presented identical self-instruction tasks in elementary geometry. The students with previous teaching machine experience were considered the low novelty effect group and the group that had been taught algebra by a regular teacher were considered the high novelty effect group.

The two groups were matched on mental ability and arithmetic achievement. The null hypothesis was tested by the use of gain scores between the initial geometry test and two others administered later in the spring semester. Measures were taken on March 6 and on June 4. It was expected that the March 6 testing might find that novelty effect very high.

The March 6 results showed that the time used and frames completed was very similar for both groups. The mean performances as represented by two groups gain scores were almost equivalent. The high novelty effect group had a mean of 21.00 and the low novelty effect group a mean of 21.80. The Mann-Whitney test showed that the results were not statistically significant. The June 4 results were all about the same as the March 6 results resulting in a general interpretation that E=C.

Rimoldi (64), in comparing methods of training in problem solving, used as a control technique provisions for equal test treatment for both experimental and control groups by introducing a substitute treatment for the experimental variable.

Two groups of students were used. The first consisted of high school freshmen (male) at St. Ignatius High School in Chicago with IQ's (Henmon-Nelson) of 120 or above whose parents gave consent. Twenty-six experimental control pairs were assembled on the basis of pretest and IQ. The second group consisted of forty-eight male college freshmen with experimental-control pairs being matched on the basis of pretest.

The experimental conditions were such that students in the experimental group were given individual training sessions in problem solving. They were permitted to ask questions related to the solution, but were given no actual instructions as to the procedure to be used. Each student was asked to evaluate his performance after he solved the problem and encouraged to talk freely about the types of questions asked. The experimenter withheld any evaluation of the performance or suggestions for better approaches. A mid-testing session was given individually without the review session. Pretests and posttests were given in group form.

The authors handled the novelty problem in the following manner: "The control subjects (high school freshmen) were given the Rotter Incomplete Sentences Blank and the Rozensweig Picture Frustration Test in an attempt to eliminate at least partially the Hawthorne effect."

Summary. From a source of over three hundred and fifty abstracted research studies, sixteen studies which specified control procedures for the Hawthorne effect were identified.

Four of these studies used the procedure of including an additional control group. The CC groups were utilized under one of the following conditions: (a) the group was established after the experiment was completed, (b) the group had no knowledge of the experiment, and (c) the group received a pseudo-treatment.

Twelve studies attempted to equalize the treatment of the experimental and the control groups as a procedure to prevent the Hawthorne effect. Variations of this procedure used most often were where neither group was told of the experiment and where steps were taken to equate motivational factors of each group. Other variations of this control procedure included both groups being told they were experimental, duration of the study assumed to eliminate any novelty effect, control group given a pseudo-treatment and an experimental school being chosen for the study to avoid the novelty effect.

The experimental results reported for these studies were examined for differences in effectiveness between the various control procedures. It was not possible to detect any empirical relationship between control procedures used and the outcomes of the studies.

Study Duration and Results

Purpose. Although the Hawthorne effect has often been cited as a variable affecting the results of experimental studies, its relative influence on these results is unknown. It has been generally hypothesized that if novelty or special attention are important aspects of the Hawthorne effect, then the strength of the effect ought to decrease with time. If there is no contaminating variable such as the Hawthorne effect operating, it might be assumed that there would be little or no relationship between duration and results. Thus, if a relationship were observed, it might be due to the contamination by the Hawthorne effect.

It has generally been assumed in educational literature that the Hawthorne effect is temporary, wearing off as the novelty of new materials and excitement of the experimentation declines with continued exposure. An early reference to the probable wearing-off of a novelty effect was made by McCall (52).

> Though evidence on this question is meager, there is some reason to believe that the mere process of experimenting with new methods or materials of instruction attracts such attention to the traits in question as to cause unconscious concentration; both on the part of teacher and pupil upon progress in these traits. As a result, it is supposed that a large but artificial growth, and that this artificial effort will evaporate if the novel methods or materials were used term after term. If each succeeding term shows a flagging of effort and an elimination or reduction of superiority, the existence of such ephemeral effort may be assumed. (52, p. 67).

Later Brownell (9) warns against the danger of assuming that the superiority of a new method is proven by the results of experimentation.

> For example, the very novelty of a new system of instruction may make it attractive to teachers and learners alike, thus giving it a special advantage, and perhaps only a temporary advantage over the rival, traditional system of instruction. (9, p. 60).

More recent references on the Hawthorne effect state that its influence should be expected to decrease with time. For example, Thorndike (79) stated that this influence should wear off with the excitement of being part of an experiment.

> This /Hawthorne effect/ is the effect that stems merely from being part of an experiment. The suggestion is that it is less important what one does to an experimental group than that one is doing

> something. Just being in an experiment, receiving special attention, and being at the center of the stage appears to influence performance, presumably through increasing motivation and effort. The effect may be expected to be temporary, and to evaporate as the excitement of being part of an experiment wears off. (79, p. 31).

Other authors feel that the influence of the Hawthorne effect should wear off with the novelty or newness of the new program. They suggest that experimental studies should then be long enough to allow this novelty to wear off. Borg (6) feels that this period should be at least over one year.

> The influence of Hawthorne effect can be expected to decrease in the school situation as the novelty of the new method wears off, and therefore, studies extending over a period of two or three years can be relied upon somewhat more in evaluating the effectiveness of a new technique. (6, p. 338).

Lee (45) also suggested that a longer period of time would reduce the influence of the Hawthorne effect. He felt this period should be several years.

> It might be suggested, therefore, that in setting up an evaluation program, a length of time sufficient for the 'newness' of the program to wear off would be highly desirable. There are no hard-and-fast rules as to the length of time that should be allowed, but a good working rule-of-thumb would probably be three to five years. (45, p. 191).

Another reference indicating that time might be related to the influence of the Hawthorne effect was made by Pella (57). He does not mention duration, *per se*, but he does feel that teachers and pupils must be used to the new method.

> Teachers employing the newer methods or tools must become proficient in their use....Pupils must become accustomed to the different technique or tool so that a "halo" or "Hawthorne" effect is eliminated. (57, p. 398).

Assuming this common belief to be true; namely, that the Hawthorne effect heightens achievement and that its influence declines with duration, one would expect that more short studies than long ones would result in the experimental groups exceeding the comparison group. The longer the study, the more often the comparison group should be equal to or exceed the experimental group in achievement. However, if the pattern would show the opposite result such that long studies more frequently than short studies result in experimental groups exceeding the control, some other explanation would have to be given. Either the influence of the Hawthorne effect is not that which is generally assumed, but is cumulative and builds upon duration, or the pattern is caused by some other variable. The purpose of this section is to present the results of an investigation aimed at answering the question, Can different results be identified through the analysis of completed studies...on the basis of the duration of the experiment?

Procedure. The analysis reported was based on the sample of research reports which had been abstracted as described earlier. As noted in Chapter II, these abstracts had been divided according to major areas of subject matter, facilitative devices, or general teaching and learning methods. Of the several major areas, six had a sufficient number of abstracts to indicate that individual analysis might be possible to show whether separate patterns can be obtained in studies within one area. The major areas concerned are mathematics, social studies, science, reading in the subject matter areas, training devices, and audio-visual communication.

The first step in the analysis was the construction of two-way tables relating the duration of the study to the results for each major area of the total group of studies abstracted. One of the first problems encountered in the construction of the two-way tables was the need to make a decision regarding the time intervals to use. As there was no special guide indicating what these should be, each major area was handled separately and the studies in that area were divided into several different time intervals that seemed to divide the studies into logical groups. Finally, an attempt was made to select a division of duration for all areas which would not obscure differences in any category so that one common division could be utilized. The intervals of time finally selected were: (a) less than two weeks, (b) two weeks to one month, (c) one month to three months, (d) three months to one semester, (e) one semester to one school year and (f) more than one school year. It was felt that a broader breakdown could easily be made more narrow by combining categories.

Study results were categorized simply as the experimental group exceeding (E>C), equalling (E=C), or not doing as well (E<C) as the comparison or control groups. Since several studies

reported more than one comparison, a decision was made to tabulate each specified result. Thus the number in each cell of the tables presented below does not necessarily indicate a separate study. In some cases, a reported study was placed in more than one square. For example, in some cases where the studies were over a school year in length, the results obtained at the end of the first year were reported in addition to the results obtained after the second, third, or fourth school year. In some studies, some part of the population was involved for only part of the total time of the study while the remainder was involved for the total period. In such cases, the study would be recorded for the proper result under each time interval for which results were reported. Further, in some studies the investigation was made on several different grade levels. In those cases where results were reported for each grade level, each result was reported. Finally, in some studies totally different results were reached for different hypotheses or with different types of instrumentation. For example, one mathematics study reported the SMSG experimental group excelling in an SMSG test, while the comparison group excelled on the conventional test. Also one reading study reported the experimental group excelling in rate, while the comparison group excelled in comprehension. In such cases, the diverging conclusions were treated separately. A final example is that in which there was more than one experimental group and one exceeded the comparison group and the other did not. Then each comparison of two groups was also treated independently. Thus the total number does not represent the total number of studies but rather the total number of different conclusions reported.

Results. The analyses of the six areas mentioned were combined to form a total number of results. Table 14 presents the data regarding study duration in relation to reported results for the total group of results.

It should be noted that the studies reporting no time duration are not included in the totals nor in the percentages calculated. This data accumulated from all the studies abstracted shows a total of 303 outcomes from the studies reporting duration, while 53 additional did not report duration. It can be seen that almost all the studies reported the experimental group exceeding or doing as well as the comparison group. Very few comparisons (only about 10 percent) showed the on-going teaching method exceeding the experimental group.

In general, the greatest proportion (40 percent) of studies were over one semester to one school year in length. An analysis of these studies showed that most of them were intended as experiments lasting one school year, although the

Table 14.--Duration versus Results of Comparative Studies Abstracted

Duration / Results	Less than 2 wks.		2 wks. to 1 mo.		1 mo. to 3 mos.		3 mos. to 1 sem.		1 sem. to 1 sch. yr.		More than 1 sch. yr.		Total		Not Reported
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	
E>C	8	6	10	7	17	12	24	17	68	48	15	11	142	47	29
E=C	9	7	11	8	27	20	36	28	39	30	8	6	130	43	17
E<C	1	3	2	6	7	23	4	13	15	48	2	6	31	10	7
Total	18	6	23	8	51	17	64	21	122	40	25	8	303	100	53

actual experimental period may have been somewhat less. According to the notion that the Hawthorne effect is a novelty effect, the superiority of the experimental group should decrease with time. The data presented in Table 14 do not reflect such a trend. When studies that had a duration of greater than one semester are combined, 56.4 percent (83 of 147) reported results of E>C while 11.6 percent (17 of 147) showed results where E<C. Of the studies less than a semester in duration, 38 percent (59 of 156) reported results of E>C and 9 percent (14 of 156) showed E<C thus revealing less significant differences for studies of relatively short duration with the experimental group showing less superiority in these studies than in the longer studies.

These data seem to show a trend opposite to the general opinion concerning the Hawthorne effect that the superiority of the experimental group should decrease with time. Of the studies reporting E>C, proportionately more had a duration of greater than one semester (59 percent of the number of E>C conclusions) than did studies reporting E=C (36 percent of the number of E=C conclusions). With an almost comparable number of conclusions in each of the two categories overall, there were more E=C conclusions in each duration interval of one semester or less in length. At the same time, there were proportionately more E>C conclusions in the duration intervals over one semester in length. The studies in which the comparison group exceeded the experimental group report 54 percent (17 of 31) with a duration of over one semester and 46 percent (14 of 31) for duration of less than one semester.

The pattern of duration versus results can perhaps be seen more clearly in Figure 4. Here again it is apparent that the data does not show any general pattern in regard to the relation of results to duration of the study.

Individual Areas. In addition to the total number of studies reported above, analyses were made of the six specific areas in which 19 or more conclusions had a reported duration. Areas omitted reported a duration for less than 15 of the conclusions.

Table 15 represents the data obtained from the studies in mathematics.

These data reflect some of the trends noted for the total group. Most of the studies here also gave results where the experimental group exceeded the comparison group (E>C, 51 percent; E=C, 37 percent; and E<C, 12 percent). Also, the largest number of studies (49 percent) had a duration over one

Figure 4.--Percentage of Studies versus Duration

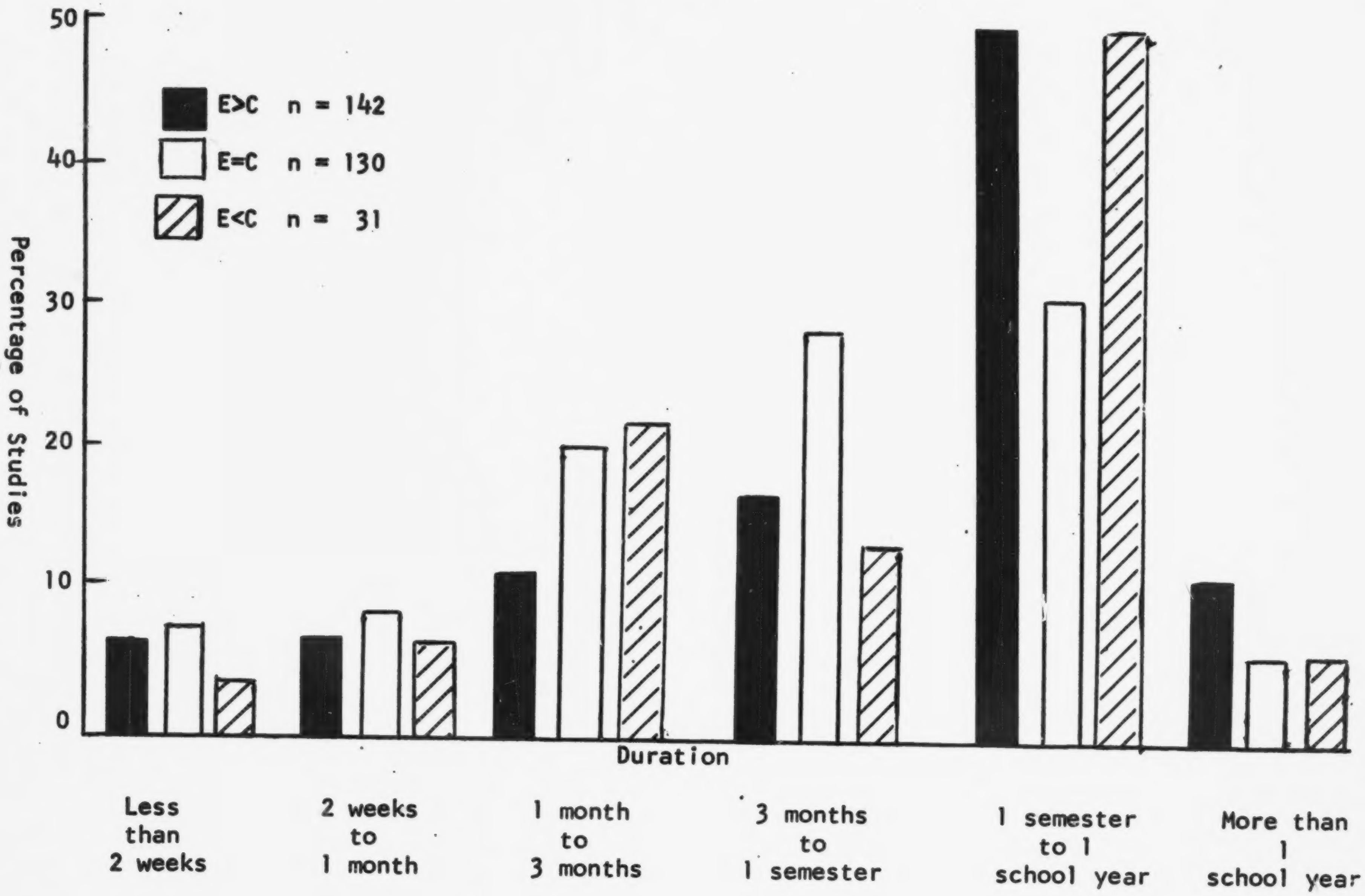


Table 15.--Duration versus Results: Mathematics

Duration / Results	Less than 2 wks.	2 wks. to 1 mo.	1 mo. to 3 mos.	3 mos. to 1 sem.	1 sem. to 1 sch. yr.	More than 1 sch. yr.	Total	Not Reported
E>C	3	4	3	5	12	2	29	2
E=C	2	2	2	1	11	3	21	2
E<C	0	1	0	0	5	1	7	1
Total	5	7	5	6	28	6	57	5

semester to one year in length. The mathematics area tended to have a slightly larger percent lasting under one month (21 percent) than did the studies in all areas combined (14 percent).

The general trend between duration and results shown for the total table is, however, not reflected in the mathematics studies. Here, the number of studies reporting durations of one semester or less show more results where E>C is greater than those resulting in E=C. The studies lasting over one semester in length result had about as many E>C conclusions as E=C conclusions. Percentage wise, about 48 percent (14 of 29) of the E>C conclusions had a duration over one semester, while about 67 percent (14 of 21) of E=C conclusions, and 86 percent (6 of 7) of E<C conclusions reported a similar duration. The trend shown by mathematics, therefore, appeared to be opposite to the general trend.

The data for the area of reading are presented in Table 16. Here again the majority of studies showed the experimental group exceeding the comparison group with most of the studies being over one semester to one school year in length. There were, however, only few studies (5 percent) less than one month in length. Another significant point is that all studies abstracted in reading reported the duration.

In general, the relationship between the results E>C and E=C and duration reflects the trend shown by the total group. However, the point at which the number of E>C conclusions exceeded the number of E=C conclusions came at one duration interval earlier (3 months to one semester). When the studies were again categorized up to the point of one semester, 65

Table 16.--Duration versus Results: Reading

Duration Results	Less than 2 wks.	2 wks. to 1 mo.	1 mo. to 3 mos.	3 mos. to 1 sem.	1 sem. to 1 sch. yr.	More than 1 sch. yr.	Total	Not Reported
E>C	0	0	2	5	11	2	20	0
E=C	2	0	3	3	7	1	16	0
E<C	0	0	1	0	2	0	3	0
Total	2	0	6	8	20	3	39	0

percent of E>C conclusions had a duration of over one semester while 50 percent of E=C conclusions had a duration of over one semester. There were only three E<C conclusions, two of which had a duration of over one semester.

The data obtained for the area of science are presented in Table 17. In this area, the number of E>C conclusions equalled the number of E=C conclusions. Again, the largest proportion of studies was in the over one semester to one school year duration, although this interval represented less than one-third of the total number of studies and almost as many studies were in the interval of one month to three months.

Table 17.--Duration versus Results: Science

Duration Results	Less than 2 wks.	2 wks. to 1 mo.	1 mo. to 3 mos.	3 mos. to 1 sem.	1 sem. to 1 sch. yr.	More than 1 sch. yr.	Total	Not Reported
E>C	0	2	3	2	6	1	14	2
E=C	1	0	5	5	3	0	14	2
E<C	0	1	1	0	1	0	3	0
Total	1	3	9	7	10	1	31	4

The studies in science also show about the same trend noted for the total group with 50 percent of the E>C conclusions and only 21 percent of the E=C conclusions resulting from studies over one semester in length. There were only three studies resulting in a E<C conclusion with only one of these being of a duration longer than one semester.

The data obtained from the comparative studies in social studies are presented in Table 18. This area had the smallest number of studies and a considerable portion of these (30 percent) did not report study duration. There were no studies reporting a duration of less than one month and only one reporting a length of more than one school year. Of the total, about half were one semester or under and half were one semester or over. The majority of results reported showed E>C conclusions.

Table 18.--Duration versus Results: Social Studies

Duration / Results	Less than 2 wks.	2 wks. to 1 mo.	1 mo. to 3 mos.	3 mos. to 1 sem.	1 sem. to 1 sch. yr.	More than 1 sch. yr.	Total	Not Reported
E>C	0	0	2	2	7	1	12	6
E=C	0	0	1	1	2	0	4	2
E<C	0	0	1	2	0	0	3	0
Total	0	0	4	5	9	1	19	8

For this area, there were three times as many studies reporting E>C as those reporting E=C conclusions. The number reporting E>C in each time period was greater than that reporting E=C conclusions. However, 67 percent of the E>C conclusions had durations of over one semester, while only 50 percent of the E=C and none of the E<C conclusions had durations of more than one semester. This again contributes to the general trend shown for the total group.

The data accumulated from the studies in Audio-visual Communication are summarized in Table 19. In this area, the number of studies resulting in no differences (E=C) made up the largest proportion, although the E>C category (38 percent) is still relatively large and the E<C category small (11

Table 19.--Duration versus Results: Audio-visual Communication

Duration / Results	Less than 2 wks.	2 wks. to 1 mo.	1 mo. to 3 mos.	3 mos. to 1 sem.	1 sem. to 1 sch. yr.	More than 1 sch. yr.	Total	Not Reported
E>C	1	3	4	3	14	5	30	14
E=C	1	4	8	15	10	2	40	5
E<C	0	0	2	1	6	0	9	5
Total	2	7	14	19	30	7	79	24

percent). Again the greatest number of E>C conclusions (47 percent) fell into the one semester to one school year duration category. The number of studies not reporting duration represented 23 percent of the total number of conclusions.

Again the pattern of duration versus results reflects that presented for the total group. The duration categories of one semester or under contained more conclusions of E=C, while those over one semester in duration contained more E>C conclusions. Only 30 percent of the E=C conclusions were over one semester in length, while 63 percent of the E>C and 67 percent of the E<C conclusions were of this duration.

The data from the studies concerning training devices are shown in Table 20. Here, the majority of conclusions were E=C

Table 20.--Duration versus Results: Training Devices

Duration / Results	Less than 2 wks.	2 wks. to 1 mo.	1 mo. to 3 mos.	3 mos. to 1 sem.	1 sem. to 1 sch. yr.	More than 1 sch. yr.	Total	Not Reported
E>C	4	1	1	2	4	0	12	1
E=C	3	4	4	2	2	1	16	0
E<C	1	0	0	0	1	0	2	0
Total	8	5	5	4	7	1	30	1

as in the A-V area. E>C conclusions accounted for 40 percent of the total number reporting duration, while E=C conclusions accounted for 53 percent, and E<C conclusions for 7 percent. Here the majority of the studies were short, with 43 percent of the total number of conclusions having a study duration of under one month. Only one study, representing 3 percent of the conclusions, did not report the duration.

Since the distribution in this area was somewhat different, the general pattern of duration versus results was again determined by percentages. Studies lasting over one semester accounted for 33 percent of the E>C conclusions, 12.5 percent of the E=C conclusions, and 50 percent of the 2 E<C conclusions. Again, the general trend was the same as that for the total group.

A summary of percentages calculated from the preceding six tables showing the relationship between duration and results of the study is presented in Table 21 when categorized as over one semester in length. The line labeled All Conclusions gives the percentage of studies in each of the areas reporting a duration greater than one semester. If there was no relationship between study duration and results, one might expect similar percentage of the studies within each area reaching each of the three different conclusions.

Table 21.--Percentage of Studies Which Were Over One Semester in Duration with Reported Results

Areas / Results	Math	Reading	Science	Social Studies	Audio-visual Comm.	Training Devices	All Areas
E>C	48	65	50	67	63	33	58
E=C	67	50	21	50	30	12.5	36
E<C	86	67	33	0	67	50	55
All Conclusions	60	59	35	53	47	27	49

Upon examining the table it can be seen that in each of the areas presented, except for mathematics, the percentage of E>C conclusion is greater than the percentage for the total number in each area as represented by the All Conclusions line.

The inverse is true for the E=C conclusions. No definite pattern can be determined for the E<C conclusions although the tendency is for the percentage of studies to exceed the percentage of the total number. In most cases, however, the small number represented by the percentage would tend to cast doubt on any significant tendency the percentages may show. In general, the longer studies tended to have more significant differences than short studies. The only area in which this appeared not to be so was mathematics.

The All Conclusions line in Table 21 gives some indication of the relative lengths of studies in different areas of research. The studies in mathematics revealed the highest percentage (60 percent) with a duration of over one semester while training devices studies tended to be the shortest with only 27 percent over one semester in length.

A summary of the number and percentage of studies abstracted which did not report duration is shown in Table 22.

Table 22.--Number and Percent of Studies Not Reporting Duration

Area / Concls.	Math		Reading		Science		Social Studies		Audio-visual Comm.		Training Devices		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
E>C	2	6	0	0	2	12.5	6	33	14	32	1	8	29	17
E=C	2	9	0	0	2	12.5	2	33	5	9	0	0	17	12
E<C	1	12.5	0	0	0	0	0	0	5	36	0	0	7	18
All Conclusions	5	8	0	0	4	11	8	30	24	23	1	3	53	15

The areas of reading and training devices had the lowest percentage, while social studies and audio-visual communication had the largest percentages (30 and 23 percent) of studies not reporting duration. Except for these two areas, there did not seem to be large differences between the three types of conclusions for that group not reporting duration.

Summary. The purpose of this section was to answer the question, Can different results be identified through the analysis of completed studies on the basis of the duration of the experiment? It was hypothesized that the superiority of an experimental group should decrease as the duration of a study increases if a Hawthorne effect were operating. The results here, while not tested for statistical significance, show a trend in the opposite direction. The longer studies more often resulted in significant differences, usually with the experimental group being superior.

Mechanical Changes and Study Results

Purpose. The second half of question 4 of prime objective A was concerned with the mechanistic nature of the experimental and control group settings as these might relate to the results of comparative educational studies.

The rationale for suspecting a possible relationship between these two variables is derived from common usages or definitions of the Hawthorne effect. It is often stated that the changes accompanying the introduction of an experimental treatment alone are sufficient to create the phenomenon of concern in this paper. For purposes of this study, the mechanistic nature of an experimental or control group setting refers to changes introduced into classrooms through the experimental procedure which might serve as cues to the subjects alerting them to the experiment, creating novelty effects, or causing a feeling of special attention. If a relationship between the nature and degree of mechanical changes in a group of comparative studies and the results of these same studies could be established, the resulting information might serve as a guide for future research activities in their desire to control the Hawthorne effect.

Procedure. The analysis of the mechanical nature of the experimental settings of comparative studies was conducted using a random sample of 53 of the over 300 educational studies which had been previously abstracted. An assumption was made that such a sample would be representative of the total number of studies abstracted, which in turn were assumed to be a representative sample of studies comparing new methods to on-going methods and are reported in educational literature.

One major procedural step in the collection of data was the definition and categorization of mechanical changes within the comparative studies. A mechanical change was operationally defined as any change in classroom procedure or instruction which was introduced by the experiment but which would not be done normally in that classroom. Only those activities which were clearly mentioned were eligible for inclusion as changes; those which could only be inferred were excluded. Changes were located as being in the experimental group, the comparison group, or both groups.

Consideration was given to the idea that not only the number but also the types of changes introduced might be useful in the analysis. Accordingly, classification categories were established by analyzing a sample of mechanical changes identified in the comparative studies. The final categories were established after a small group of studies had been analyzed and their mechanical changes identified and categorized.

Two major categories were established. The first distinguished between intentional and incidental cues. Intentional cues would be those activities or changes which were an integral part of the variables being manipulated and compared. An example of an intentional cue would be the use of SMSG materials in a comparison to conventional materials. Thus, they were deliberate and obvious changes due to the experiment. Incidental cues were changes that were not intended as variables. An example of this would be the use of pretests which were not a normal part of the school curriculum. These changes may be either necessary in the design of the experiment or inadvertently introduced when adapting the experiment to the school situation.

The second categorization distinguished changes into various types as to their nature. Six sub-categories of changes were identified as physical, personnel, materials, equipment, verbal, and instructional procedures. Physical changes were such activities as moving to a different classroom or a different building for instruction. These types of changes would obviously be most important in studies on the elementary level, but may assume some importance on other levels also. Personnel changes involved such things as having a different teacher, being in a team-teaching situation, or being supervised by a non-teaching assistant with no regular instructor. Material changes involved the use of a different textbook or instructional materials, expecially materials in experimental form, changes in the kind or frequency of the use of visual aids (when no special equipment is involved) or the use of unusual pretests or special inventories (as attitude, personality, etc.). Equipment changes included the use of teaching machines, television, unusual increase in the use of visual aid, equipment, etc. Verbal cues involved such changes as deliberately informing the subjects of the experiment, giving them special encouragement, or alerting them with unusual instructions. Instructional procedures changes were the use of individualized or group instruction (if not employed previously), a change from primarily lecture instruction to primarily discussion, and changing from a laboratory approach to a demonstration approach in a science class.

Analysis of the studies was initiated by identification and categorization of changes as reported in the abstract. Since some of the abstracts appeared incomplete, the original reports from which the abstracts were taken were re-examined as necessary. When this latter action proved almost fruitless in the identification of additional changes, the remaining abstracts were assumed to be as complete as the report from which they were taken. After changes were identified, they were categorized as to type. At least two members of the project staff were involved in the identification and categorization of the changes noted in each study.

After the changes were identified and categorized, the study was identified according to the results obtained. Then the data from the different studies were combined so that the final tabulation could be analyzed to determine whether any pattern existed.

A summary of the data is presented in Table 23. A few words of explanation concerning the tabulation procedure will help in interpreting the table. As can be easily seen, the total number of comparison groups is less than the sum obtained from the number of comparison groups listed under the three columns: E>C, E=C, E<C. In addition, the total number of experimental groups is greater than the total number of studies. This was due in all cases to the use of multiple experimental groups in a study. Each experimental group is reported only once but in those cases to the use of multiple experimental groups in a study. Each experimental group is reported only once but in those cases where one experimental group exceeded the comparison group and another group did not, then the comparison group is listed more than once in each column--E>C, E=C, E<C--whichever applied. However, if the same result in a particular study was obtained with each experimental group relative to the comparison group, then the comparison group was listed only once in the appropriate column while each experimental group was listed separately.

A survey of Table 23 seems to reveal no definite trends relating the numbers or types of changes to the results reported. However, a general comment may be made concerning the results. As might be expected, the majority of the mechanical changes were made in the experimental groups and, of these, most were identified as intentional cues. This would seem to imply that the majority of the cues identified were directly related to the

Table 23.--Mechanical Changes in Comparative Studies as Related to Results

E>C

CHANGES	EXPERIMENTAL GROUP		COMPARISON GROUP	
	Intentional Cues	Incidental Cues	Intentional Cues	Incidental Cues
Physical	0	2	0	0
Personnel	1	0	0	0
Materials	7	3	1	3
Equipment	11	2	1	2
Verbal Cueing	1	0	1	0
Instructional Procedures	13	0	0	0
Column Total	33	7	3	5
Group Total	28		24	
Number of Studies - 24				

E=C

CHANGES	EXPERIMENTAL GROUP		COMPARISON GROUP	
	Intentional Cues	Incidental Cues	Intentional Cues	Incidental Cues
Physical	0	0	0	1
Personnel	1	0	0	0
Materials	4	7	0	3
Equipment	21	0	0	1
Verbal Cueing	1	3	0	1
Instructional Procedures	14	2	0	1
Column Total	41	12	0	7
Group Total	31		25	
Number of Studies - 25				

E<C

CHANGES	EXPERIMENTAL GROUP		COMPARISON GROUP	
	Intentional Cues	Incidental Cues	Intentional Cues	Incidental Cues
Physical	0	0	0	0
Personnel	1	2	0	0
Materials	3	2	1	1
Equipment	4	3	1	2
Verbal Cueing	1	0	1	0
Instructional Procedures	5	0	0	0
Column Total	14	7	3	3
Group Total	10		9	
Number of Studies - 9				

instructional variable. The most frequent types of changes were in materials, equipment, and instructional procedures.

This analysis did not show a trend in any direction which would support a hypothesis that the results of comparative studies are related to the mechanistic nature of the experimental and comparison group settings.

The purpose of this chapter was to present the results of the literature phase of the project as they pertain to those objectives and questions relating to conceptualization of the Hawthorne effect, and suggested and actual procedures employed to control for its operation in an experimental situation, plus possible relationships between duration of a study and the nature and type of mechanical changes as both of these relate to the results provided in completed research studies.

Chapter V following presents a discussion of the results from both the field research and literature analysis phases of the project along with the conclusions drawn from the results. The chapter is concluded with a presentation of some implications of the total project effort for educational research.

Chapter V

Discussion, Conclusions, and Implications

Introduction

Chapters III and IV have presented the results from the field research and the literature analysis phases conducted to secure information to answer the questions set forth in the objectives. This chapter presents (a) a discussion of the results obtained from the two major phases of the project, (b) a statement of the conclusions which appear to be justified on the basis of the results, and finally (c) possible implications of the conclusions with regard to the role of the Hawthorne effect in educational research. Separate sections within this chapter are devoted to each of these topics.

Discussion

Since there were two major research tasks conducted in the form of a field research and a literature analysis, it seems desirable that a discussion of the results deal separately with each of these tasks. Consequently, the discussion has been divided into two parts for clarity of presentation.

Field Research Results. To review briefly, the essential purpose of the field research was to determine the feasibility of establishing a classroom experimental situation in which direct and indirect cues would be introduced to determine if such cues had any subsequent effect on student achievement as measured by gains over a period of time. Concurrent with this feasibility study was an attempt to determine if the presence of such cues operated in a manner such that student awareness of experimental participation would be heightened and such awareness would in turn be related to the gains in achievement. Combined with these two purposes of the field research was a third purpose which focused upon the question of whether or not extending the duration or time length of an investigation was an appropriate control technique for a possible Hawthorne effect.

The general results from the field research indicated that the various combinations of direct and indirect cues did not operate in a manner such that achievement was significantly effected. Further, no significant relationships were observed between awareness and achievement gains. Extending the study for a two year period did not produce results showing significant differences in achievement at time periods selected to represent extended study duration.

A review of the general experimental design employed plus an examination of specific procedures suggests several possibilities as to why the results reported above may have occurred.

One distinct possibility is that the variable under concern, the Hawthorne effect, simply does not exist as a variable of sufficient potency to be significantly influential on study results. On the other hand, it is possible that the phenomenon may exist but that it spreads equally over all treatment conditions, and thus its influence in an experimental investigation is minimized if not altogether eliminated. If either of these interpretations possess validity, educational researchers could proceed to disregard the possible operation of the phenomenon in their investigations and accept significant differences as being basically due to the independent variable (s) introduced as part of the experimental design.

Interpretation that it does not exist, at least as evidenced by the results from the present investigation, does, however, have one fundamental limitation. It is, in effect, generalizing from the single case. That is, the experimental situation here represents a single effort to study the role of direct and indirect cues. Replication of the study would have to be made in order to provide a larger sample of studies on which to establish a more reliable statement that the Hawthorne effect is not present in curriculum research investigations.

Returning to the specific procedures, it may have been entirely possible that factors constituting the several treatment combinations applied to the several groups under the experimental design were not sufficiently potent enough in their novelty, newness, or difference from normal school routine that they would generate an increased motivation for achievement. This situation may have arisen because of the desire to maintain within the experimental design a pattern of instructional research typical of many classroom experiments yet at the same time attempting to avoid the specific concept under study. It might well be that the desire to capitalize upon the natural or normal curriculum break at the end of third grade to introduce a new curriculum was not unique enough since students were perhaps experiencing or had experienced similar conditions in other areas of the curriculum. In order to accomplish the objectives of field research, it may have been wiser to introduce variables or factors which may have been more unique and/or novel than those that were actually introduced but such a step would require prior knowledge of the potency associated with such variables.

Another explanation might center around the criterion instrumentation, the STEP Mathematics test, which may not have been a sufficiently valid measure of the actual instruction so that it would reflect achievement gains initiated by the cues that were employed. The criterion instrumentation problem is, of course, not unique to a study of this type since it is present in almost all attempts to evaluate a new curriculum. While the awareness scale did reveal differential awareness between the various treatment groups, this instrument might be limited in its contribution by reflecting only a cognitive awareness of participation and not any possible affectional influence such knowledge may have had upon achievement.

Another possible explanation might center around the schedule of measurements. In order to maintain an experimental situation very similar to traditional classroom methodology studies, normally scheduled school testing periods were used as much as possible. It may well be that whatever effect is generated by the new or unusual treatment combinations, it appears early; that is, within the first few weeks or months of an experimental situation. In the present study, such an effect may have occurred and then leveled off during the first year and also over the two years thus accounting for the absence of any significant differences at these time periods. Should this condition have actually occurred in the field research, it would not have been possible to ascertain it since no attempt was made to secure more frequent assessments than those indicated in the procedures section. It was felt that any procedure set up to do so might have in and of itself confounded the experimental situation.

Another possible explanation is that in a study running over two years time, even with the introduction of treatment combinations as presented above, it would appear exceedingly difficult to separate out normal gains in achievement due to student growth and development from those attributable to the treatment combinations. It seems safe to assume that whatever is done to students some growth and development will take place leading to achievement gains over a one or two year period. To introduce selected variables with the idea of generating awareness and thus causing greater than normal achievement might seem presumptuous in the sense that one would experience difficulty in partialing out that growth due to experimental conditions and that due to normal conditions. A close approximation to the solution of this problem might be made by reference to the traditional control group achievement but even here the assumption is made that achievement gains in the control group are the result only of normal growth and not the result of possible indirect or direct cues provided by the experimental arrangements. It is interesting to note that in the present study, the control treatment did indicate some awareness of experimental participation albeit at the lowest end of the awareness score continuum for the eight treatment groups.

Another consideration that should be set forth in interpreting the results of the field research phase relates to the initial differences observed on the criterion measure. The impact of this situation on any resulting effects due to application of the treatment conditions is not known. It would be difficult to speculate on how the field experiment might have gone had there not been the initial interaction observed between the Teacher and Method factors. As noted, adjustments were made by using as covariates initial measures in reading and mathematics. Upon completion of the field research phase, an analysis was made of the several covariate scores using only the subjects who had remained over the two years of the study. The results of this analysis are presented in Table 24. Inspection of this table reveals that not only was there the Teacher-Method interaction noted originally but there was now a significant interaction between these two factors for the Word Knowledge and the Comprehension subtests of the Metropolitan Reading Test. Further, significant F-ratios were observed for the Teacher and Informing main effects on the Comprehension subtest. These results indicate that the initial class differences were not only present but that these differences probably continued throughout the study as was noted by the significant interactions between Teacher-Method and Teacher-Informing noted in the analyses of true gains reported in the Field Research Results chapter.

The results obtained from the study of awareness-gains relationships run counter to what amounts to a common belief regarding the role of subject awareness in a classroom experiment. This belief usually states that subjects who exhibit awareness of their participation in an experiment tend to have larger gains than might be normal (i.e., superior to that of control groups) under an implicit assumption that awareness produces motivation which in turn produces greater achievement. Several explanations might be advanced with regard to the observed relationships which do not support this belief.

First, the correlation presented for the end of the two year period was obtained by relating the mean awareness score for a treatment combination to the mean gain score for the same treatment condition using only two measures per treatment to calculate the means. A question might be asked, Were the treatment combinations themselves correlative to the awareness

Table 24.--Summary Tables for Analyses of Variance for Selected Pretest Variables

Source	d.f.	STEP I Mathematics		Metropolitan Reading			
				Word Knowledge		Comprehension	
		MS	F	MS	F	MS	F
Teacher	1	0.867	.211	0.595	.167	20.342	6.574*
Method	1	0.308	.075	6.900	1.943	1.979	.640
Informing	1	2.620	.636	14.019	3.948	25.606	8.275*
Teacher-Method	1	79.580	19.324**	287.0064	80.8264**	220.292	71.191**
Teacher-Informing	1	0.027	.066	18.916	5.327	12.3158	3.980
Method-Informing	1	0.124	.030	0.507	.143	0.204	.066
Teacher-Method-Informing	1	0.681	.165	11.150	3.140	1.562	.505
Error	7	4.118	--	3.551	--	3.094	--
Total	14	8.0739	--	25.9964	--	21.7115	--

*F(1,7;.05)= 5.59 **F(1,7;.01)= 12.2

measures? One answer to this question is provided in Appendix B which presents information on the development of the awareness instrument. In the development phase of this instrument, significant differences were observed between the average awareness scores for each treatment with those treatments having more changes (new text, different teacher,) receiving higher awareness scores. It would be somewhat difficult to establish a meaningful correlation coefficient between the awareness scores and the treatment conditions since the latter were not considered to be ordered in any particular manner. One might order them in terms of the number of changes associated with each treatment but that would have the effect of placing different cues under a single descriptive label (e.g., one change). If the treatments were classified by the number of changes, then one might find himself putting together such treatments as Conventional Mathematics--Different Teacher--Not Told with Conventional Mathematics--Regular Teacher--Told since only one change was made within each treatment. For this reason, questions about a possible relationship between the treatment conditions and awareness cannot be adequately resolved by the evidence obtained during the development phase.

A second explanation might center around the idea that those students under a treatment condition who showed a relatively small gain over the time period involved tended to be those with initial high achievement scores and thus not able to exhibit much gain. It would seem likely that such students would also be the brighter or more intelligent students. Again, evidence presented with regard to the development of the original instrument indicates that there was no distinct relationship between the awareness scores for treatments and a measure of mental ability. It is quite possible that there might be some relation between gain and mental ability, but an examination of the realtionships between gain and mental ability by treatment groups as reported in Appendix B revealed no meaningful relationships.

A further possible explanation of the results might be that the Hawthorne effect does exist within students and classes, but it might well be considered as a variable of insufficient potency to cause any real differences. This might possibly explain why the relationships between awareness and gains varied between classroom groups within treatments. That is, within a given single treatment combination one classroom group might have a positive relationship between gain and awareness while the second classroom might have a negative relationship. Should this assumption have validity, researchers could release themselves from needless worry about possible contamination of research by a Hawthorne effect. This idea is only a suggestion at this stage of the investigation and should be examined further.

It is also quite possible that if the measures of gain and awareness had been obtained earlier in the study rather than at the end of the two year period, the correlation might have been higher. This would be particularly true under a hypothesis that any Hawthorne effect which does appear in an experimental situation appears rather early and thus has its most influence during the initial phases of an experimental situation. The effect is then subsequently reduced over time. In view of the conditions under which this investigation was conducted, it is not possible to provide an answer to this question. Concurrent with this idea of early measurement is a possible subject attrition interpretation. It is quite possible that the elimination of certain subjects because of incomplete data at the termination of the two year period may have had the effect of reducing the variability in both the gain and awareness measures and thereby reducing the correlation coefficients.

Still another possibility is that awareness and gain are two variables which do not have any fundamental relationship between them. Awareness of experimentation may be one human attribute and achievement gain another, and to posit a relationship between them in the sense that awareness is an independent variable and achievement a dependent variable is not reasonable. The evidence accumulated in this study is not sufficient, however, to provide an answer to this question.

A final explanation might be that the measure of awareness developed within this project simply did not measure or assess awareness of experimentation in a manner which would allow possible motivational awareness to be ascertained. In brief, we simply did not measure the right thing.

As noted, the field research phase was conducted over a two year period in order to secure evidence relative to a commonly presented suggestion that extending the duration of an investigation is a reasonable technique for controlling or even eliminating the possible influence of novelty, newness, or enthusiasm as a causative factor in producing increased levels of achievement. The results from this phase of the study revealed no significant differences in achievement scores both at the end of a one year period or at the end of a two year period.

The results of the investigation on extended duration of an investigation as a control procedure are perhaps most easily explained in view of the fact that no significant results were obtained during the course of the entire field research. Had there been significant differences in achievement at the end of the first year followed by nonsignificant differences at the end of two years, such results could have been interpreted as being supportive of the hypothesis that such a technique would be a valid control procedure.

Several of the explanations discussed in connection with the experimental design, such as instrumentation and potency of independent variables, also have applicability in explaining the results of this aspect of the field research. The findings presented are, however, limited because no measurements were taken within the first month or two to see if there were early significant differences in achievement between treatment combinations which were then reduced or eliminated at the end of one or two years. Some researchers have indicated that if there is any effect due to novelty or similar factors, it probably appears early and then is gradually diminished. Some thought was given to securing data early in the field research, that is at some time during the first half-year, to test this hypothesis. It was felt, however, that in view of the major concept being investigated it would be inappropriate to introduce the needed measurements early in the experiment since such instrumentation itself might contaminate the field research design which was created so as to be similar in nature to many traditional classroom experimental situations. Consequently, such a procedure was not included in the research. It is suggested that other researchers might approach this problem by using a time-series analysis technique within informal or non-experimental situations as contrasted to formal experimental situations where the introduction of measurements might influence negatively the internal validity of the design.

Literature Analysis Phase. The literature analysis phase was undertaken to determine if information regarding the role of the Hawthorne effect in educational experimental designs could be secured through an examination of completed research studies. More specifically, the focus was to be on determining if experimental-control group differences could be explained on the basis of whether or not and how an investigator might have attempted to control for the phenomenon, the duration of the study, and the nature and type of mechanical changes introduced.

Before discussing the results, it appears appropriate to comment upon the use of published literature as a source of data for answering research questions based upon the experiences in this project.

The use of completed research studies as data for the investigation of research questions and objectives is a questionable technique for four reasons.

First, a rather limited amount of information about the objectives, procedures, and data tend to be reported in the published literature. Project staff members preparing abstracts often encountered a lack of sufficient information to know exactly what was done during the course of an investigation.

Second, the studies that were abstracted represented a large number of diverse experiments involving totally different populations, many different designs, and a wide variety of independent, dependent, and even contaminating variables. Thus, it was very difficult to make meaningful comparisons between and among studies.

Third, a lack of a sufficient number of experimental studies dealing with varying approaches to instruction and curriculum to answer some of the questions was also encountered. One question dealt with the idea of trying to find out if experimental groups exceeded control groups no matter what the independent variable was within the same design structure. For example, if one study had "word sight" as the new method in contrast to "phonics" as the new method and "word sight" as the old method, would the experimental group still be superior. The pursuance of this objective was not possible because there were just not enough published studies in the available sample to make such a comparison. There were many experimental studies comparing a "word sight" to "phonics" but not many in the other direction.

Fourth, the published literature contains an imbalance, if it can be called that, with regard to the number of studies published presenting significant differences in favor of experimental groups as contrasted to the number of published studies showing no differences or differences in favor of the control group. This observation has been made previously by many persons, both in and out of educational research. The cause of this situation has, however, not been fully explored. Editorial policies, reluctance to report non-significant results, and other factors may be causative. Regardless of the cause, such a situation presents limitations if one anticipates using the published literature as a source of data in examining possible factors other than planned independent variables which might explain significant or nonsignificant differences between experimental and control groups.

Several comments seem appropriate with regard to the more specific findings reported in the results chapter for the literature analysis. The comments presented below relate to the three general questions investigated as part of this phase of the total project.

One of the major findings of the literature analysis was that there is definitely a lack of agreement regarding the nature of the Hawthorne effect concept. This is evidenced by the several definitions and/or interpretations of the Hawthorne effect cited in that section of the results dealing with definitions and suggested controls. This situation is very analagous to a situation described by Anthony (3) relative to topics which might be included under the role of operations researchers.

> ...the flitting around from topic to topic makes it impossible to find a unifying concept that will serve to delineate the area so that it can be distinguished from other areas. Lacking such a concept, operations researchers are at best Edisons - geniuses in their own rights, but not pointing the way toward future development - as contrasted with Faradays, who develop a conceptual scheme that stimulates further progress. (3, p. 91).

An attempt was made early in this paper to develop a conceptualization which might provide a focus for further work in this area. It may turn out in the long run to be a truly inappropriate approach. On the other hand, turning to the literature to find a conceptualization of the Hawthorne effect upon which there is agreement appears not to be a fruitful approach at least as based upon the definitions presented in this report.

The problem of ascertaining meanings and definitions of the concept was complicated by the fact that the concept generally is not indexed in any systematic manner so that it can be easily located. Many of the definitions, as well as suggested and detected control procedures, were discovered by accident upon reviewing a completed study or in a discussion of experimental procedures. This fugitive nature of the concept made the task of securing definitions, meanings, or interpretations of the concept a much more difficult task than was originally anticipated. In view of this situation, the suggestion is made that the term "Hawthorne effect" become a major retrieval term within the ERIC system of the U. S. Office of Education. Should this be done, perhaps other researchers can, in subsequent years, probably derive a larger sample of meanings and thus develop a more meaningful and useful concept.

One of the main questions investigated was efficacy of the suggested and actual control procedures used by persons who indicated some awareness of the Hawthorne effect problem. It is recognized at the start that the results presented here are restrictive in the sense that the control categories used in the review of such procedures were but one method of grouping. Another point of view might have approached the categorization of control methods into those that operated from the general basic design of the experiment and those that introduced procedural activities specifically formulated to control for the Hawthorne effect. Studies included in the first group would be those using a control-control group in some manner and those where both groups were told that they were the experimental group or both told they were the control group. The distinction between the above is that in all these cases the control was executed through the fundamental reorganization of the experiment and applied at the subject level as opposed to application at the teacher level.

Within the sample of studies examined, the employment of a control-control group provided an opportunity to derive from measurement an indication of the effectiveness of the control whereas the remainder had no such opportunity because of a lack of a baseline measure. The second sample of studies that instituted specific procedural activities arising from the unique conditions of their study also had no opportunity through measurement to determine whether the control functioned as assumed. The point being made here is that even though such control methods were employed, the investigators did not attempt to study the effectiveness of the procedures used. The failure of the researcher to take such an action should not be interpreted as a weakness of the study. The presentation of data relative to the effectiveness of control methods employed would have made the task of determining the efficiency of the method in the present investigation a much easier task.

A question might be asked, What has the above analysis revealed to persons interested in controlling for the Hawthorne effect? One major point is that there should not be a simple reliance on the strength of assumption alone that any control method employed is functioning. More empirical evidence is needed to support a researcher's position that the employed control procedure is working.

The value of the use of a control-control group has been clearly demonstrated by those investigators whose studies were included in the results section. The direction of attention to procedural activities specified for each study, while not quantitatively reassuring when used alone, cannot be denied a distinct necessity in careful research. Some examples of the kinds of consideration made in studies with a variety of structures included in this analysis suggest factors to be aware of in future studies. It would seem that a combination of the use of a control-control group and the introduction of activities to equalize the influence of the experimental situation between the experimental and control groups is perhaps the most comprehensive approach possible, at least as indicated by the results reported with regard to actual control procedures used in completed research studies.

A discussion of attempts to establish reliable procedures for controlling the Hawthorne effect clearly exposes the circularity of the problem and makes generalizations from the results presented above unreliable for several reasons. One major limitation is the multitude of sites where the Hawthorne effect might operate in the most meticulously designed experiment. Its imprecise definition, as noted above, permits it to exist in various degrees in studies despite calculated efforts to protect against it. A second major limitation is the small number of studies found where the investigators were sufficiently aware of the possibility of profound influences on their research results that the Hawthorne effect exerts that they actually preplanned active control for it. Further, the studies that did include controls for the Hawthorne effect were so diverse in all variables that evaluation of a portion of each intimately integrated control system does not reasonably yield generalizations. Thus, a limited number of examples make it difficult to derive more definitive rules for effective control and the lack of definitive rules makes it discouraging for the investigator to attempt control.

Returning to the original question posed about the relationship between reported results and the method of control employed, the results presented reveal a composite of an overwhelming number of both planned and chance events in each investigation. Any relationship between the control methods employed and experimental results, which may exist, was not exposed by the procedures employed in the present investigation.

That aspect of the literature analysis dealing with the duration of a study and its relation to results was undertaken to secure evidence regarding the usefulness of this technique as control procedure. The general finding reported above was that the longer studies (more than a year in length) showed a larger proportion of differences favoring the experimental group than was the case for shorter time periods. Several possible explanations for this finding might be advanced.

One is that a Hawthorne effect, if operating, may actually be cumulative thus increasing differences between experimental and control groups with time instead of diminishing such differences.

Another possible explanation is that one school year is not long enough time for any possible Hawthorne effect to diminish. This idea would seem to be in accord with the views of Borg and Lee noted earlier who recommend three to five years as the period of time in which such effects should wear off. However, this idea is not indicated by the studies in our sample having durations over one school year. These studies followed the same trend as that shown by studies of one school year in length. The smaller number of shorter studies showing significant differences might then be simply due to a lack of sufficient time to develop such differences. One weakness of the analysis which makes it hard to accept or reject this explanation was the relatively small number of experiments lasting more than one school year.

As noted, the data gathered to answer the question was limited in terms of the number of studies with extended time duration. It may be that with increased federal support for extensive curriculum projects lasting closer to the period of time suggested by Borg and Lee, additional evidence might be forthcoming which would provide a more definitive answer than that obtained in the present investigation.

The results presented with regard to mechanical changes and study results revealed no distinctive trend which would support a hypotheses that the results of comparative studies are related to the mechanistic nature of the experimental and comparison group settings.

This analysis is limited by the common restriction that such findings depend to a great extent upon whether or not such changes can be truly identified from the study report as well as the validity of the categories established into which the several changes were classified. Since in many of the studies the number of changes were not clearly specified, many subjective decisions had to be made with regard to the nature and type of change introduced. The classification scheme employed was established both ad hoc and a posteriori. It is certainly within reason that the original investigators might not consider the actions that they took during the course of the investigation to be classifiable as mechanical changes. It may well have been that many of the activities regarded here as changes were actually only secondary activities from the view point of the original investigator.

The results presented are also limited by the fact that the data presented was derived using a small number of studies selected from the total number of available abstracts. A decision was made that a sampling from the total pool of abstracts would provide sufficient evidence to answer the question. As with any sampling plan, another set of abstracts might have provided a different set of data leading to a different finding.

Conclusions

The statement of objectives in Chapter I indicated that the overall goal of the project was to assemble a body of information about the Hawthorne effect which might provide a basis for the subsequent development of a theory to explain the phenomenon as it relates to educational experimentation. It was not the intent of the project to either initiate or complete the development of such a theory. It was anticipated that the total project effort would provide some degree of information which would allow educational researchers to make decisions with regard to the role of the effect in experimental designs during the interim period until such a theory is developed and tested.

As has been stated by many writers on the decision-making process, information relative to possible alternatives is a necessary and prerequisite step to final choice by those having responsible administrative and managerial positions in a research situation. This project was oriented to trying to assemble and organize in some systematic manner the present state of information with regard to the Hawthorne effect in educational research so that better decisions might be made. Much has been written and said about the role of the Hawthorne effect in educational research but very little effort has been made to integrate the currently available information. As Anthony has pointed out with regard to the topics of planning and controlling in the management context.

> Isolated experiences and discrete bits of knowledge are not very useful. When organized into some kind of pattern, however, the pieces often illuminate one another; the whole becomes greater than the sum of its parts. The very act of organizing may show that the framework itself is defective. Nevertheless, the framework will have served a useful purpose if it prepares for a better one. (3, p. iii).

The essence of this project was to organize the existent information about the Hawthorne effect into some type of framework. It is recognized that in time additional information will be forthcoming which, without doubt, will have a substantial influence upon further clarification of the role of the Hawthorne effect in educational research.

Since the basic purpose of the project was largely to secure information for subsequent use, conclusions in the traditional hypothesis testing sense seem inappropriate. Some specific hypotheses were established and tested, particularly in the field research phase, but such tests were conducted as part of the data collection activity relevant to a particular question or objective and not as general overall hypotheses for the total project.

The statement of conclusions presented in this section with regard to the role of the Hawthorne effect in educational research are qualified by a rather essential point, namely the conceptualizations and definitions of the term or concept itself. To conduct an investigation of the phenomenon as represented by the field research phase of the project, an operational definition was developed which focused upon the subject's awareness of experimental participation as outlined in the conceptualization provided in Chapter I. Concurrent investigation of the phenomenon as represented by the literature analysis phase of the project had to accept the definitions or meanings provided for the term by those researchers discussing it as documented in Chapter IV.

As a consequence of the above situation, the information accumulated with respect to the phenomenon reflects some diversity rather a consensus of meanings. The questions established to guide the study required, however, that answers be presented in a general rather than a specific manner and employing principally an inductive approach. The conclusions presented below, therefore, are in the form of general answers to the questions utilizing evidence and data gathered from the two separate phases of the project rather than in terms of specific answers restricted to specific definitions and meanings.

The validity of such a procedure is perhaps open to question but this action was taken in view of the stated overall purpose of the study which was to secure information regarding the present status of the Hawthorne effect concept in educational research.

Three general or major questions were stated originally along with a series of specific questions derived from each of the major questions. Each major question is presented below along with its related specific question followed by that data or evidence which appears to have relevancy in answering the question. A final statement is then made relative to each specific question and the general question.

A. Can the Hawthorne effect be identified as the variable accounting for all or a major portion of the difference or lack of difference noted between experimental and control groups in selected educational research studies?[1]

A-1. In completed educational research studies of similar content and approach, are there differences in relative control and experimental group achievement results and study conclusions when varying approaches have been used by the investigations to control for the Hawthorne effect.

Data gathered from the literature analysis relevant to this question leaves the final answer as being somewhat inconclusive. This answer is made on the basis of the fact that it was difficult to locate a sufficient number of research studies of similar content and approach to provide necessary data and the absence of any reasonable empirical evidence presented by the investigator that the control procedures actually employed were effective in controlling for the effect. The data that was accumulated with regard to this question, however, does suggest a tentative answer that one cannot explain differences between study results on the basis of the procedures employed to control for the possible presence of the Hawthorne effect. This tentative answer is suggested primarily on the basis of the fact that most researchers approved to have worked from an assumptive basis only that the procedures actually employed to control for the effect did so rather than from any empirical evidence as to their effectiveness.

---

[1] The question numbered A-2 in the statement of objective in Chapter I which read, Are there groups of completed educational research studies which, when taken in combination, have produced similar results for different methodological approaches when the common element of the studies has been the construction of the design. That is, for example, do reading methods studies show more consistency in results on the basis of whether the adaptation (i.e., phonics vs. word-sight) is experimental or control than on the basis of what the experimental method is? is not included in the statement of conclusions because insufficient evidence was available to answer the question in terms of its demands. That is, it was not possible to find studies in which the experimental design was the same but where there was a reversal of methodologies of concern as experimental variables and control variables.

A-2. Can an experimental situation be contrived within which the results or findings of "methods studies" will vary as a consequence of the way in which the Hawthorne effect is controlled or not controlled in the experiment?

The evidence relevant to this question is drawn mainly from the field research phase of the project. Here an attempt was made to create a traditional classroom type experimental situation utilizing a series of direct and indirect cues which were conceived as producing and/or controlling for the Hawthorne effect.

No significant differences between the several treatment combinations were observed at the end of two years of the investigation. It should be recognized that attempts to control for the Hawthorne effect as it is traditionally interpreted were not applied since the aim of the field experiment was to determine if the effect, defined here as awareness of experimental participation, could be created within elementary school students. To the extent students involved in different combinations of cues revealed differential awareness, one could state that a control for the Hawthorne effect was established. At the same time, the data reveal that such awareness was not related to subsequent achievement by the students.

In view of the results obtained, the answer to this question would be negative to the degree that special efforts to control an experimental situation failed to produce substantial relationships between the awareness of participation and achievement results.

A-3. Can different results be identified either through the analysis of completed studies or the construction of an experimental situation on the basis _of the duration of the experiment or the mechanistic nature of the experimental and control group settings?_

Data relevant to this question can be derived from both the field research and literature analysis phases of the project.

The results from the field research phase indicated that extending the study over a two year period did not produce any substantial change in achievement gains in the sense that treatment group differences were modified at the end of two years as opposed to the end of the first year of the study. Achievement results at the end of the first year were not significantly different between the various treatment combinations of direct and indirect cues.

Data from the literature analysis with regard to the duration of the study and experimental results revealed a relationship contrary to what seems to be a logical position that novelty and similar effects should diminish over time. The evidence accumulated here indicated that the longer the study was conducted the more significant differences favoring experimental over control groups as contrasted to equality between the two groups or the control exceeding experimental group were observed. Such evidence, however, has to be tempered by the fact that more studies of the first type were reported in the literature than were studies showing no differences or the control group exceeding the experimental group for the same time period.

Evidence from the field research phase contributes little to the answer but the literature analysis does. It indicates that the answer to the question inquiring about differences in study results in terms of the duration of the study would be negative in the sense that there appeared to be no reduction of experimental group superiority to control groups over time, but actually an increase which would be contrary to an expectation that the Hawthorne effect could be used to account for initial superiority followed by diminishing of this superiority.

The introduction of mechanical changes in the form of indirect and direct cues in the field research phase as represented by a new method, change in teachers, and informing and not-informing students of their participation appeared to have no effect upon achievement results as measured by gains in mathematics achievement over a one and two year span. Some interactions between the factors introduced were noted but these appeared to be a function of initial differences between the participating classroom groups. Based upon the evidence from the field research phase of the project, the answer to that part of the question asking about mechanical changes is largely negative. Evidence from the literature analysis with regard to mechanical changes revealed no significant pattern of relationship which would permit a statement that the more such changes were introduced into either experimental or control groups the more likely there would be a significant difference between the groups.

The general answer to that part of the question dealing with mechanical changes is largely negative in the sense that it does not appear possible to explain differences between studies with regard to experimental results on the basis of a variety of mechanical changes being introduced which might create a Hawthorne effect.

On the basis of the answers provided to the series of specific questions above, it would seem justifiable to state that the answer to the general Question A would be negative. It appears unlikely that one can employ a Hawthorne effect concept to explain differences or the lack of differences between experimental and control groups in educational research studies in so far as the variables commonly believed to generate the effect such as direct and indirect cues, the duration of a study, and mechanical changes introduced in an experiment are considered to be of sufficient potency to produce the effect.

B. Can the component elements of the so-called Hawthorne effect be identified?

B-1. Is subject awareness of participation in an experiment sufficient to create differences in achievement results in control and experimental groups?

Data relevant to this question is drawn from the field research phase and particularly from that part of the investigation dealing with the possible relationship between gains in achievement and awareness of experimental participation. Even though subjects in the several treatment groups responded differentially on a scale designed to measure awareness of participation, the resulting awareness measures were not significantly related to achievement. Data from the teacher interviews also provides some additional evidence in the form of observations that several students recognized something unusual going on, but that such interest was short-lived and transitory in nature and thus probably had no substantial effect on performance.

The evidence accumulated from this activity leads to a negative answer to the question asking if subject awareness of experimental participation is sufficient to create differences between experimental and control groups.

B-2. Are the relationships in question B-1 of the same direction and magnitude where indirect cues are used as contrasted with situations using direct cues?

Evidence relative to this question can be derived from the data on the gain-awareness correlations for each of the several treatments which involved different combinations of direct and indirect cues. This evidence reveals no definitive pattern which would indicate that gain-awareness relationships are of a similar direction and magnitude within those treatment combinations having similar direct and/or indirect cues. Further evidence can be secured from the study of mechanical changes where an attempt was made to classify the changes as intentional (direct) or incidental (indirect). In this analysis, no definitive relationship appeared to exist between either type of change and experimental results indicating that such changes appeared to have cue value for the participating subjects. Based upon this evidence, the answer to this specific question is negative.

The answer to the general question B, asking if the component elements of the Hawthorne effect can be identified, appears on the basis of the answers to questions B-1 and B-2 to be negative. Awareness appears not to be a useful total component. Apparently there appears to be other components which yet need to be identified before the question can be answered more definitively.

C. From an analysis of questions A and B, can the relative efficiency and effectiveness of control mechanisms be inferred?

C-1. If no overt or purposeful cues are provided to the control group, can the investigator assume that the Hawthorne effect is not operating within the control group?

Evidence to answer this question is drawn from the field research phase. If the conceptualization of the Hawthorne effect proposed in this report has validity, the answer to this question would have to be in the negative. Measurements of awareness of experimental participation secured from the control group (Conventional Mathematics--Regular Teacher--Not Told), while lower in magnitude than those from other treatment combinations as noted both in Chapter III and Appendix B, nevertheless was sufficiently high on the scale of awareness to indicate that some knowledge of special arrangements was present.

C-2. To what extent are placebo treatments necessary as control mechanisms?

The evidence available to answer this question comes principally from the literature analysis phase of the project. In some of the cited research studies where actual control procedures were built into the experimental design, such procedures took the form of providing to the control group a set of conditions designed to equalize the attention paid to the experimental group or the novelty effect presumably created within this group by the introduction of the experimental variable. Such procedures as were employed, however, did not appear to be substantiated on the basis of empirical data that they did serve as intended or were even perceived by the control group participants as placebo treatments. The answer to this question appears at the present time to be inconclusive. To establish an answer, evidence should be collected which would substantiate the premise that a treatment designed to function as a placebo did in effect have this function. No such evidence was accumulated as part of the present investigation.

C-3. Can extended duration of the experiment be used as self-control for the Hawthorne effect?

A definitive answer to this question is not possible since no data or evidence was gathered in either phase of the project revealing early presence of a Hawthorne effect which in turn modified by subsequent extension of the experiment. The correlation between final awareness scores and gains at the end of the first year of the field research was not substantially different from that obtained at the end of two years. These data are limited, however, since the awareness score was secured at the end of the two year period. They do indicate that had there been any viable modification of an early Hawthorne effect, as herein defined, by extension of the study over a two year period, then such a procedure might be considered as a useful technique. A more specific answer to this question is, however, largely inconclusive based upon the data presently available.

An answer to the general question C, inquiring about the relative effectiveness and efficiency of control mechanisms from the evidence providing answers to Questions A and B, would be principally in the negative. This answer is based upon the data relating to control procedures actually employed but not validated on the basis of empirical evidence that they were effective and the data relating to duration of the study and the nature and type of mechanical changes, both of which were shown not to be related to experimental results in a manner such that one utilize these variables to explain possible differences between experimental and control groups. It would appear that at the present time no efficient and effective control mechanisms for the Hawthorne effect based upon empirical evidence are identifiable. Until such evidence is secured, educational researchers will apparently have to continue working from an assumption that the control procedures they do employ are effective.

The results of the present study would appear to indicate, however, that awareness of experimental participation, at least on the part of elementary school students, is not the contaminating variable that it has been traditionally assumed. Hence, extensive efforts to control for this situation may not be fundamentally necessary. The related literature cited in Chapter I regarding subject awareness in experimental situations at the adult levels, as described by Orne and Rosenthal, would tend to indicate that the above statement might be invalid for experimental situations at the secondary school and college levels.

Implications

The previous sections of this chapter have presented a discussion of the results obtained from the two principal phases of the project along with a series of conclusions in the form of answers to questions which the data and evidence secured during the course of the project seem to support. The purpose of this section is to set forth what appear to be the principal implications of the total research effort with regard to the nature and function of the Hawthorne effect concept in educational research. Three principal implications appear to be justifiable in terms of the results and conclusions presented above.

The first implication is that perhaps much of what has been written up to the present time about the nature, operation, and control of the Hawthorne effect in educational research projects comparing different instructional and curriculum approaches appears to have been generated largely on the basis of intuition and logic rather than upon any empirical basis. The present investigation would indicate that many of the commonly held ideas and beliefs about the role of the Hawthorne effect lack sufficient validity at the present time to warrant their serious consideration by researchers in the planning and operational development of their projects as well as in the interpretation of results. For example, a rather common belief is that subject awareness of experimental participation will have a contaminating effect on the project activities in the sense that such awareness would generate positive results. The evidence gathered in this project would tend to indicate that students could be aware of an experimental situation, particularly at the elementary school level, and still not jeopardize the results of the study in terms of the independent variables being investigated. This is not to say that subject awareness might not jeopardize the results of projects conducted at other higher levels of education since evidence has been accumulated by researchers concerned with the social-psychology of experimentation to indicate that there are facilitative effects at such levels. The age level at which these latter groups of studies are conducted would probably be most similar to the age level at which the original Hawthorne studies were conducted indicating perhaps that any discussion of the Hawthorne effect have to be considered in terms of an age variable. In brief, it might appear that many individuals in education, researchers included, have taken a phenomena identified as being existent at the adult level and hypothesized its existence at non-adult levels and controlled for it but without empirical evidence to substantiate either the hypothesis or the controls.

Correlated with this implication is that researchers in the past have tended to assume that the introduction into an experimental situation of a wide variety of changes, particularly those associated with an experimental independent variable, have created a novelty, attention, or enthusiasm condition which results in a positive effect upon the results of the study. The present investigation would indicate that such a concern need not be overemphasized since the evidence presented here indicated that no meaningful relationship was established between number and types of changes and experimental results. If the relationship presented here does have validity, it would suggest that researchers could feel free to introduce with the independent variable as many changes as seemed desirable and feel that the results would be jeopardized.

Concurrent also with this same implication is a common belief that if a Hawthorne effect is created, it appears early in the experiment and therefore extending the study for a longer period of time would tend to minimize or even eliminate any such effect.

The results presented here indicate that this suggestion does not appear to be a valid technique or procedure since no reasonable evidence was gathered that there were not more significant differences between experimental and control groups in the results of short run studies as contrasted to studies conducted over a longer period of time. In fact, the reverse situation seems to hold.

In addition to the above, the present investigation also suggests that in those cases where researchers have indicated a general awareness of the Hawthorne effect problem some rather elaborate control procedures have been adopted which reflect or involve a rather large amount of effort on the part of the research project staff and/or the participants. Such efforts often include, for example, a rather extensive training program designed to make sure that teachers in a control group are given equal treatment to the experimental group teachers so that the Hawthorne effect is equalized. The present study would indicate that any such efforts have been initiated largely on the basis of an assumption that such treatments are necessary for equalization purposes, and that they are effective. The present study was unable to ascertain that such assumptions can be substantiated through empirical evidence.

The employment of such procedures tends to indicate further that researchers seem to expect that when they carry out a research investigation the Hawthorne effect is going to operate, and therefore need to control for it. To suggest that they do otherwise or not even be concerned with its possible presence would almost seem to be unscientific. Any researcher who has been well trained in the scientific method would be concerned with possible sources of uncontrolled variation. The situation here might be somewhat analogous to believing the existence of a particular type of effect to be encountered in a space flight[2] on an intuitive basis and then developing elaborate control procedures with no empirical evidence to establish the presence of the assumed effect. Continued operation of space projects on this basis would lead to such project engineers as being called unscientific. It is more likely that research would be conducted to justify the existence of the assumed effect, and if found to exist, then there would be need to develop elaborate controls. This thought leads to what appears to be a second major implication of the findings from the present study.

---

[2] The early ideas about the possible consequences of "crashing the sound barrier" by aircraft would be a similar situation.

3

The second major implication is that the Hawthorne effect appears to have been placed in the position of being guilty with efforts then directed to prove its innocence. Such a situation is not only contrary to the spirit of constitutional law but is also contrary to the nature of science. It is quite possible that the rather frequent negative answers to the questions raised in the present study could indicate that the Hawthorne effect does not exist in as an operationally definable construct. That is, a construct may have been established but which is not producible or measurable in any meaningful operational way. If this situation is true, the implication is that any concerns about the effect at the present time could be disregarded by researchers in their creation of research projects and in interpreting the results.

Another idea closely related to the above is that perhaps the Hawthorne effect might best be considered as a random variable associated with individual subjects as opposed to groups. It may be that individual subject behavior operates in such a manner that any subsequent effect on experimental results appears to be operationally spread over all treatment conditions. It has not been specifically stated but researchers seem to have acted as if the Hawthorne effect is a group phenomenon. The assumption here is that all the members of an experimental group, because of their participation, exhibit a similar degree of "enthusiasm." If so, then perhaps the point made by Miles as cited in Chapter I would have validity. He pointed out that defining the Hawthorne effect in terms of awareness of experimental participation by individual subjects was an inappropriate approach for the present investigation. He further suggested that changes in the social sub-system in which the experiment is placed should be the focus. The present study did focus on individual awareness and not upon the social sub-system in which the experiment was contained. Here again, however, is the suggestion noted earlier that adult interactions within a social sub-system have an impact upon the results of production, and therefore, such a condition would exist in other less sophisticated or non-adult social systems in which experiments are placed. Statements of this type lead to what appears to be a third implication which is developed in the form of a recommendation for further research.

The third implication relates to an observation made during the course of the project and reported in Chapter IV that was a lack of agreement with regard to the meaning or definition of the Hawthorne effect among educational researchers. It would appear from this situation that a rather strong implication of the present research would be that educational researchers should devote some time and energy to arriving at an operational definition of the Hawthorne effect which has sufficient focus and delimitation that research studies could be conducted to determine its precise role in an experimental situation. Control procedures and/or techniques for partialing out the results of such a phenomenon in the analysis of research investigations could then be developed.

Some progress is being made with this type of research as represented by the work of individuals engaged in investigating the social psychology of experimentation. As appears to be true in many areas of education, the needed research on a vital topic is coming from what might be called non-education fields. In this case, the evidence is coming largely from the fields of psychology and sociology. If educational researchers sincerely believe, and the principal investigator believes they do, that the Hawthorne effect is a variable causing them great concern in the conduct of their research, it would seem any time and energy be devoted to securing valid and reliable evidence with regard to the concept would be justifiable.

The present investigation represents only a start on the above topic and, without doubt, suffers from many limitations. It may produce controversy or it may be ignored. Regardless of what reaction appears, it is hoped that the publication of the study will prompt other researchers to initiate further studies of the problem so that concerns about the Hawthorne effect in educational research do not operate solely on an intuitive and assumptive basis which, up to the present time, appears to have been largely the way that the phenomenon has been handled.

Chapter VI

Summary

The Problem

The Hawthorne effect has been the subject of much discussion in the area of experimental educational research even since the phenomenon was observed, discussed, and reported in the Western Electric Company studies on the relationship between work efficiency and illumination. One of the general findings of this research was that the worker production tended to increase over time regardless of the nature and type of changes in illumination that were introduced into the work environment. After consideration of other possible causes, the consequent production increase was said to be caused by the fact that the workers felt they were the objects of special attention from management and this condition motivated them to produce more than might be normally expected.

This same basic idea of special attention has been often cited to explain the results of many research studies conducted in the field of education to determine if a new teaching method or curricular improvement might be superior to an existing, on-going, or current patterns of teaching or curriculum. Many researchers have explained the presence of significant superiority of the new over the old method, or even the lack of such differences, as being due to the possible operation of the phenomenon which has been referred to as the Hawthorne effect. The superiority of the new procedure over the on-going is said to be a result of the subject's perception that he is the object of special attention and/or the novelty created by a change from the current classroom situation.

Even though the Hawthorne effect has been often utilized to explain such differential results, there is a lack of information about the precise nature and operation of the phenomenon as it relates to experimental designs in educational research.

The general objective of the present study was to build up a body of information about the Hawthorne effect which would provide a basis for the subsequent development of a theory to explain the effect. Under this general objective, the work of the project was directed to securing data and evidence to provide answers for three general questions each of which had a correlative series of more specific questions. The three general questions were (A) Can the Hawthorne effect be identified as the variable accounting for all or a major portion of the difference or lack of difference noted between experimental and control groups in selected educational research studies? (B) Can the component elements of the so-called Hawthorne effect be identified? and (C) From analysis of A and B above, can the relative efficiency and effectiveness of control mechanisms be inferred?

Procedure

The methods of inquiry employed to secure the necessary data and evidence to answer the above series of questions consisted of two principal operations, a field research and a literature analysis phase.

The field research phase consisted of developing a factorial experimental design which could be considered as being representative of traditional types of comparative method research studies in the field of education. The principal purpose of the experimental situation was to study the role of one direct and two indirect cues as independent variables on a dependent variable of subsequent student achievement. The direct cue consisted of information provided to elementary school students that they were in an experiment. The indirect cues consisted of the introduction of both a new curriculum in the form of the SMSG mathematics and the introduction of a different teacher for the mathematics period of instruction. Each of the independent variables or experimental factors had two levels: SMSG and Conventional mathematics, regular and different teacher, and told and not told of experimental participation. Under this design, 8 different combinations were established and randomly assigned to 16 elementary school classroom groups (2 classes per treatment) selected from the Columbus, Ohio Public School System. The field research was conducted over a two year period using fourth grade students initially and continuing on with the same students during the fifth grade. The purpose of conducting the study over the two year period was to secure data evidence regarding the oft made suggestion that extending a study is a possible control technique for the Hawthorne effect.

Gains in mathematics achievement at the end of one year and the two year period were used as the principal criterion measures. A paper-pencil inventory to secure an assessment of student awareness of participation in the experimental situation was developed during the course of the field research and administered at the end of the experimental phase.

Measures of awareness were correlated with subsequent gains in achievement at the end of one year and over the two year period to see what relationship might exist. An analysis was also made of the differences between mathematics achievement for the several treatment combinations at the end of one and two year periods to study the suggestion of extending study duration as a control technique.

The literature analysis phase of the project consisted of establishing a sample of abstracted published research studies which compared a new to a traditional or on-going pattern of instruction or curriculum in order to determine if the Hawthorne effect could be used to explain possible differences in results between such studies as well as to examine the efficiency of various control procedures employed by the investigators. The literature analysis was concerned also with securing data and evidence to answer questions about possible relationship between (a) study results and the nature and number of mechanical changes introduced into the experimental situation, and (b) the duration of time a study was conducted.

The data for this part of the study consisted of approximately 350 completed research studies selected from professional journals, final reports of funded research projects, and similar sources on the basis of criteria established for relevance of the report to the purposes of the present study. The research reports abstracted were drawn largely from the decade 1953 to 1963 but studies outside of this period were included if considered appropriate.

Results

The specific results or findings of the study are presented below in two sections with the data and evidence from each phase being used subsequently to answer the several general and specific questions which, although presented as answers to questions, can be considered as conclusions of the study.

The data from the field research phase of the project provided the following results:

1. No significant differences were noted with regard to the influence of the main effects, as represented by the direct and indirect cues, on student gains in mathematics either at the end of the first, second, or both years of the study. Some significant interactions were noted at the end of both one and two years of the study but such interactions were considered to be attributable to initial differences between the classroom groups to which the treatment combinations were assigned.

2. No significant relationships were observed between measured awareness of participation in an experimental situation and gains in mathematics achievement at the end of the first, second, and full two year period.

3. No significant differences were obtained between achievement at the end of one year as contrasted to achievement obtained at the end of the full two years of the study.

The data from the literature analysis phase of the study provided the following results:

1. Individuals writing about the Hawthorne effect concept were not in agreement upon the meaning or definition of the concept. Consequently, they varied with regard to suggested control procedures for the effect and who should be the recipients, teachers or students, of the controls suggested.

2. The actual control procedures employed by researchers who indicated some awareness of the effect did not appear to be related to experimental results in any systematic way that would permit a statement to be made with regard to the effectiveness of the several control procedures employed or the ability to differentiate between similar studies presenting different results.

3. Both actual and suggested control procedures appeared to be justified on an intuitive or logical basis rather than upon any empirical evidence that such procedures did in fact control for the presence or operation of the effect, regardless of the definition or meaning given to the concept by the researcher.

4. The analysis of experimental study results as related to the duration of a time study conducted indicated that studies conducted over a period of one or more years showed a greater proportion of differences in favor of the experimental group than was the case for results showing experimental equal or less than the control during the same time period. A similar result was observed for studies conducted over shorter durations of time.

5. The analysis of the nature and number of mechanical changes introduced into an experimental situation in relation to experimental results revealed that no differentiation could be made between study results by noting the degree of changes introduced.

Conclusions

Based upon the data and evidence obtained from the two major activities of the project leading to the results presented above, answers to the three general questions were established as follows:

1. On the basis of the answers provided to the series of specific questions, it would seem justifiable to state that the answer to general Question A would be negative. It appears unlikely that one can employ a Hawthorne effect concept to explain differences or the lack of differences between experimental and control groups in educational research studies in so far as the variables commonly believed to generate the effect such as direct and indirect cues, the duration of a study, and mechanical changes introduced in an experiment are considered to be of sufficient potency to produce the effect.

2. The answer to the general Question B, asking if the component elements of the Hawthorne effect can be identified, appears on the basis of the answers to the specific questions to be negative. Awareness appears not to be a useful total component. Apparently there appears to be other components which yet need to be identified before the question can be answered more definitively.

3. An answer to the general Question C, inquiring about the relative effectiveness and efficiency of control mechanisms from the evidence providing answers to Questions A and B, would be principally in the negative. This answer is based upon the data relating to control procedures actually employed but not validated on the basis of empirical evidence that they were effective and the data relating to duration of the study and the nature and type of mechanical changes, both of which were shown not to be related to experimental results in a manner such that one utilize these variables to explain possible differences between experimental and control groups. It would appear that at the present time no efficient and effective control mechanisms for the Hawthorne effect based upon empirical evidence are identifiable. Until such evidence is secured, educational researchers will apparently have to continue working from an assumption that the control procedures they do employ are effective.

Implications

Three implications of the research reported herein appear to be derivable from the questions investigated, the data and evidence collected, the results obtained, and the conclusions drawn.

1. Researchers have established the Hawthorne effect as a variable contaminating study results when there is very little evidence or information that the effect has been operationally defined in a manner such that its role in comparative educational studies can be firmly established. In essence, the Hawthorne effect concept is being put in the position of being guilty (i.e., operational) with efforts then being directed to establish its innocence rather than being considered as innocent (i.e., nonoperational) until its guilt has been established.

2. The control procedures suggested and actually employed for those situations where researchers appear to be aware of the effect appear to have been established on the basis of intuition and logical biases rather than upon empirical evidence. Consequently, no valid generalizations can be made about the effectiveness of various control procedures.

3. Educational researchers appear to have neglected, as a fruitful field of inquiry, the area of investigation represented by the Hawthorne effect and its implications for research methodology. It is suggested that educational researchers follow the lead established in other behaviorial sciences and conduct research on the social psychology of classroom experimental situations in order to provide more meaningful and useful information and evidence about the role of the Hawthorne effect in experimental designs in educational research than exists at present.

References

1. Adams, J. Laboratory Studies of Behavior Without Awareness, *Psychological Bulletin*. Vol. 54, 1957. Pp. 151-171.

2. Adrion, Jo Ann. *The Problem of Awareness and Its Relationship to Counseling*. Unpublished Master's thesis, The Ohio State University, 1964.

3. Anthony, Robert N. *Planning and Control Systems: A Framework for Analysis*. Division of Research, Graduate School of Business Administration, Harvard University, Boston, 1965.

4. Blum, Milton L. The Hawthorne Studies, *Industrial Psychology and Its Social Foundations*. Harper and Brothers, Inc., New York, 1956.

5. Boch, R. Darrell. Multivariate Analysis of Variance of Repeated Measurements, in *Problems in Measuring Change*. Chester W. Harris (Ed.), University of Wisconsin Press, Madison, Wisconsin, 1963.

6. Borg, Walter R. *Educational Research: An Introduction*. David McKay and Company, 1963. Pp. 338-339.

7. Bowers, Norman D. & Soar, Robert S. *Evaluation of Laboratory Human Relations Training for Classroom Teachers; Final Report of Studies of Human Relations in the Teacher-Learning Process*. Cooperative Research Project No. 469, U. S. Office of Education, 1961. Pp. 142-144.

8. Brickell, Henry M. *Organizing New York State for Educational Change*. Commissioner of Education, State Education Department, Albany, New York, December 1961.

9. Brownell, W. A. A Critique of Research on Learning and Instruction in the School, *Fifteenth Yearbook of the National Society for the Study of Education, Pt. I, Graduate Study in Education*. H. B. Henry (Ed.), Chicago: University of Chicago Press, 1951. Chapter 6. P. 60.

10. Campbell, D. T., and Stanley, J. C. Experimental Designs for Research on Teaching, in *Handbook of Research on Teaching*. N. L. Gage (Ed.), Rand-McNally, Chicago, 1963. Chapter 5.

11. Clark, Harold F. *Cost and Quality in Public Education.* Syracuse University Press, Syracuse, New York, 1963.

12. Cook, Desmond L. The Hawthorne Effect in Educational Research, *Phi Delta Kappan.* Vol. 44, December 1962. Pp. 116-122.

13. Cook, Desmond L. The Impact of the 'Hawthorne Effect' in Experimental Designs in Educational Research, *Appendix A.* Cooperative Research Project No. 1757, U. S. Office of Education, 1962. P. 3.

14. Cook, Desmond L. *Program Evaluation and Review Technique: Applications in Education.* Cooperative Research Monograph No. 17, U. S. Office of Education, U. S. Government Printing Office, 1966.

15. Cook, Desmond L. *The Relationship Between Possible Hawthorne Effect Components and Stages of Experimental Investigation.* Occasional Paper 63-133, Bureau of Educational Research and Service, The Ohio State University, 1963.

16. Coon, Lewis H. *Applying PERT to a Variable Manipulating Project.* Occasional Paper 64-157, The Bureau of Educational Research and Service, The Ohio State University, 1964.

17. Cooper, James G. Effects of Different Amounts of First-Grade Oral English Instruction Upon Later Reading Progress with Chamorro-Speaking Children, *Journal of Educational Research.* Vol. 58, No. 3, November 1964. Pp. 123-127.

18. Crespi, L. P. The Interview Effect in Polling, *Public Opinion Quarterly,* Vol. 12, 1948. Pp. 99-111.

19. Cronbach, Lee. Evaluation for Course Improvement, in *New Curricula.* R. W. Heath (Ed.), Harper and Row, New York, 1964.

20. Crosby, Gwladys, Fremont, Herbert, & Mitzel, Harold. *Mathematics Individual Learning Experiment,* Title VII, Project No. 391. NDEA Act of 1958, Grant No. 7-42-1050 and 7-30-0690. (Queen's College, Flushing, New York).

21. Downing, John. *The Evaluation of i.t.a.* University of London Institute of Education. Mimeo.

22. Downing, John & Haliwell, Stanley. The Ita Reading Experiment in Britain. In Figurel, J. Allen (Ed.) *Improvement of Reading Through Classroom Practice*. International Reading Association Conference Proceedings, Vol. 9, 1964. P. 261.

23. Duncan, Roger. An Experimental Study of the Effect of Parents' Knowledge on Student Performance in SMSG Mathematics, *The Journal of Educational Research*. Vol. 58, No. 3, November 1964. Pp. 135-137.

24. Edwards, Allen L. *Techniques of Attitude Scale Construction*. New York: Appleton-Century-Crofts, Inc., 1957.

25. Entwisle, Doris R. Set-Learning in Gifted High-School Students, *School Review*. Vol. 70, 1962. Pp. 473-482.

26. Eriksen, C. W. (Ed.) *Behavior and Awareness--A Symposium of Research and Interpretation*. Duke University Press, Durham, North Carolina, 1962.

27. Forehand, Garlie A. Psychological Measurement and Research in the Teaching of English, in *Needed Research in the Teaching of English*. E. R. Steinberg (Ed.), Cooperative Research Monograph No. 11, U. S. Office of Education, 1963.

28. Hall, Charles E., and Cramer, E. M. *A General Purpose Program to Compute Multivariate Analysis of Variance on an IBM 7090 Computer*. Biometrics Laboratory, The George Washington University, Cardinal Printing Co., Arlington, Virginia.

29. Harris, Chester W. (Ed.) *Encyclopedia of Educational Research*. Third Edition, The Macmillan Company, New York, 1960.

30. Harris, Chester W. (Ed.) *Problems in Measuring Change*. University of Wisconsin Press, Madison, Wisconsin, 1963.

31. Hershberger, Wayne A. & Terry, Donald F. *Complexity of Typographical Cueing in Programmed and Conventional Texts*. American Institute for Research, Palo Alto, Office of Naval Research, Report No. 7, December 1963. Pp. 6-7, 31.

32. Herzog, E. *Some Guidelines for Evaluative Research.* Children's Bureau, Social Security Administration, U. S. Office of Health, Education, and Welfare, 1959.

33. Higgins, Conwell & Rusch, Reuben R. Remedial Teaching of Multiplication and Division: Programmed Textbook Versus Workbook - A Pilot Study, *The Arithmetic Teacher*. Vol. 42, January 1965. Pp. 32-38.

34. Homans, G. C. The Western Electric Researches, in *Fatigue of Workers: Its Relation to Industrial Production.* Reinhold Publishing Corp., New York, 1941.

35. Houston, W. Robert and DeVault, M. Vere. Mathematics In-service Education: Teacher Growth Increases Pupil Growth, *The Arithmetic Teacher*. Vol. 10, No. 5, May 1963. Pp. 243-247.

36. Hovland, I.; Lunsdaine, A. A.; and Sheffield, F. D. *Experiments on Mass Communication.* Princeton University Press, Princeton, New Jersey, 1949.

37. Kaplan, Abraham. *The Conduct of Inquiry.* Chandler Publishing Co., San Francisco, 1964.

38. Keating, Raymond F. *A Study of the Effectiveness of Language Laboratories.* Institute of Administrative Research, Teachers College, Columbia University, 1963.

39. Krasner, L. Studies of the Conditioning of Verbal Behavior, *Psychological Bulletin.* Vol. 55, 1958. Pp. 148-170.

40. Lambert, Phillip, *et al.* *Classroom Interaction, Pupil Achievement and Adjustment in Team Teaching as Compared with the Self-Contained Classroom.* Final Report, Cooperative Research Project No. 1391, U. S. Office of Education, 1964.

41. Lana, R. Pretest-Treatment Interaction Effects in Attitudinal Studies, *Psychological Bulletin.* Vol. 56, 1959. Pp. 293-300.

42. Landesberger, Harry. *Hawthorne Revisited.* Cornell University Press, Ithaca, New York, 1958.

43. Lang, Gerhard & Tohtz, Jack Erwin. *Verbal Aptitude and Effectiveness of Out-of-Class Programmed Instruction in Freshman English.* Paper presented at AERA, Chicago, February 1964.

44. Lawrence, Frederic K. Measuring Subject Awareness of Experimentation. Unpublished Master's thesis, The Ohio State University, 1965.

45. Lee, Eugene E. With All Deliberate Speed, School Science and Mathematics. Vol. 63, 1963. Pp. 190-191.

46. Likert, R. A Technique for the Measurement of Attitudes, Archives of Psychology. No. 140, 1932.

47. Lord, Frederic M. Elementary Models for Measuring Change, in Problems in Measuring Change. Chester W. Harris (Ed.), University of Wisconsin Press, Madison, Wisconsin, 1963.

48. Lord, Frederic M. The Measurement of Growth, Educational and Psychological Measurement. Vol. 16, 1956. Pp. 421-437.

49. MacCorquodale, K. and Meehl, P. E. Hypothetical Constructs and Intervening Variables, Psychological Review. Vol. 55, 1948. Pp. 95-107.

50. Matarazzo, J. D., et al. Interviewer Influence on Duration of Interviewee Speech, Journal of Verbal Learning and Verbal Behavior. Vol. 1, No. 6, May 1963. Pp. 451-458.

51. Mayo, Elton. The Hawthorne Experiment, Western Electric Company, The Human Problems of an Industrial Civilization. Division of Research, Graduate School of Business Administration, Harvard University, Boston, 1946.

52. McCall, William A. How to Experiment in Education. New York: The Macmillan Company, 1923. P. 67.

53. Miles, M. B. (Ed.) Innovation in Education. Bureau of Publications, Teachers College, Columbia University, New York, 1964.

54. Oberholtzer, Kenneth E. Letter to Desmond L. Cook re: The Effectiveness of Teaching Reading in Kindergarten, Colorado State Department of Education.

55. Orne, Martin T. On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications, American Psychologist. Vol. 17, 1962. Pp. 776-783.

56. Payette, Roland F. Summary Report of the School Mathematics Study Group Curriculum Evaluation, School Mathematics Study Group Newsletter No. 10. November 1961. Pp. 5-11.

57. Pella, Milton O. Criteria for Good Experimental Research in the Teaching of Science, Science Education. Vol. 45, December 1961. Pp. 42-51, 396-399.

58. Pella, Milton O.; Stanley, Julian; Wedemeyer, C. A.; & Wittich, W. A. The Use of the White Films in the Teaching of Physics, Science Education. Vol. 46, No. 1, February 1962. Pp. 6-21.

59. Pennock, G. A. Industrial Research at Hawthorne: An Experimental Investigation of Rest Periods, Working Conditions, and Other Influences. The Personnel Journal. Vol. 8, 1929. Pp. 296-313.

60. Popham, W. James. The Influence of Novelty Effect Upon Teaching Machine Learning. Project Report Under Research Grant 7-14-1230-155.0, U. S. Office of Education, San Francisco State College, 1962.

61. Porter, Douglas. An Application of Reinforcement Principles to Classroom Teaching: The teaching of phonemically and orthographically analyzed spelling materials by an automatic mechanical teaching device. U. S. Office of Education Contract SAE-6899, Harvard University, May 1961. Pp. 37-39.

62. Psychological Research in Education. National Academy of Science, National Research Council, Publication 643, 1958. Pp. 23-24.

63. Reicken, Henry W. A Program for Research on Experiments in Social Psychology, in Decisions, Values, and Groups. N. F. Washburne (Ed.), New York, Pergamon Press, 1962.

64. Rimoldi, H. J. A.; Fogliatto, Hermalinda M.; Haley, John V.; Reyes, Isabel O.; Erdman, James B.; & Zachariz, Rose Mary. Training in Problem Solving. Loyola Psychometric Laboratory, Publication No. 27, Loyola University, Chicago, Illinois, 1962.

65. Robinson, H. Alan. Critique of Current Research in College and Adult Reading. The National Reading Conference Yearbook. No. 13, 1963. P. 172.

66. Roethlisberger, F. J. and Dickson, William J. Management and the Worker. Harvard University Press, Cambridge, Massachusetts, 1947.

67. Rosenbloom, Paul C. Large Scale Experimentation with Mathematics Curriculum, in Research Design and Analysis. R. O. Collier and S. M. Elam (Eds.), Second Annual Phi Delta Kappa Symposium on Educational Research, 1961. Chapter 1B.

68. Rosenthal, Robert. On the Social Psychology of the Psychological Experiment: With Particular Reference to Experimenter Bias. Paper presented at the American Psychological Association Annual Meeting, New York, 1961.

69. Rosenthal, Robert. On the Social Psychology of the Psychological Experiment, The American Scientist. Vol. 51, June 1963. Pp. 268-280.

70. Rudell, Arden K. and Brown, Gerald W. In-Service Education in Arithmetic: Three Approaches, Elementary School Journal. Vol. 9, April 1964. Pp. 377-382.

71. Ryan, T. Antoinette. Testing Instructional Approaches for Increased Learning, Phi Delta Kappan, Vol. 46, No. 10, June 1965. Pp. 534-536.

72. Scriven, Michael. The Methodology of Evaluation. Publication 110, Social Science Education Consortium, Purdue University, Lafayette, Indiana, 1966.

73. Sequential Tests of Educational Progress. Cooperative Test Division, Educational Testing Service, Princeton, New Jersey, 1959.

74. Sjogren, Douglas D. Programmed Materials in High School Correspondence Courses. Cooperative Research Project No. 1534, U. S. Office of Education. P. 77.

75. Smith, Gerald R. A Picture of Research As Reflected in Proposals Submitted to the Cooperative Research Program. Mimeo. undated.

76. Snow, C. E. Research on Industrial Illumination, Tech Engineering News. Vol. 8, November 1927. Pp. 257-282.

77. Star, S. A. and Hughes, H. M. Report on the Educational Campaign: The Cincinnati Plan for the United Nations, *American Journal of Sociology*. Vol. 55, Pp. 389, 1949-50.

78. Thorndike, Edward L. & Lorge, Irving. *The Teacher's Word Book of 30,000 Words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

79. Thorndike, Robert L. *The Concepts of Over- and Underachievement*. New York: Bureau of Publications, Teachers College, Columbia University, 1963. Pp. 31-32.

80. Veatch, Jeannette. In Defense of Individualized Reading. *Elementary English*. Vol. 37, 1960. Pp. 227-234.

81. Winer, B. J. *Statistical Principles in Experimental Design*. New York: McGraw-Hill Book Company, Inc., 1962.

82. Wittich, W. A.; Pella, Milton O.; Stanley, Julian; & Wedemeyer, C. A. *The Wisconsin Physics Film Evaluation Project*. The University of Wisconsin, 1961. P. 51.

83. Wolf, Willavene; Huck, Charlotte; & King, Martha. *An Investigation of Critical Reading Ability of Elementary School Children*. Cooperative Research Project No. 2612, U. S. Office of Education, 1963.

BIBLIOGRAPHY

I. The Hawthorne Effect Concept

A. References pertaining to the original work condition investigations conducted at the Hawthorne (Chicago) Plant of the Western Electric Company.

Blum, M. L. Industrial Psychology and Its Social Foundations. New York: Harper and Brothers, 1956. Chapter 2.

Chase, Stuart. Men at Work. New York: Harcourt, Brace and Company, 1941. Chapter 2.

Guba, Egon G. Research in Internal Administration--What Do We Know? In Campbell, Ronald, & Lephan, James (Ed.), Administrative Theory As a Guide to Action. Midwest Administrative Center, The University of Chicago, 1960.

Homans, George Casper. Group Factors in Worker Productivity. Fatigue of Workers: Its Relation to Industrial Production, National Research Council of the National Academy of Sciences, Washington 25, D. C., 1961. Chapter 4, Pp. 448-460.

Landesberger, H. A. Hawthorne Revisited. Ithaca, New York: Cornell University Press, 1958. Chapter 2.

Longenecker, J. G. Principles of Management and Organizational Behavior. Charles E. Merrill Books, Inc., Columbus, 1964. Pp. 12-13, 293-295.

Mayo, E. The Human Problems of an Industrial Civilization. New York: Macmillan and Company, 1933.

Pennock, G. A. Industrial Research at Hawthorne. The Personnel Journal, Vol. 8, 1929. Pp. 296-313.

Roethlisberger, F. J., & Dickson, W. J. Management and the Worker. Harvard Business, Division of Research, Business Research Studies, No. 9, 1934.

Roethlisberger, F. J., & Dickson, W. J. Management and the Worker. Cambridge, Massachusetts: Harvard University Press, 1939.

Roethlisberger, F. J. *Management and Morale*. Cambridge, Massachusetts: Harvard University Press, 1941. Chapter 8.

B. References containing comments about the role of the "Hawthorne Effect" in educational research investigations, including comments on its control.

Adrion, Jo Ann. The Possible Role of the Hawthorne Effect in Completed Educational Research Studies. Paper presented at the 1966 Annual Meeting of the American Educational Research Association, 1966.

Barrett, H. O. Large Group Instruction. *Ontario Journal of Educational Research*, Vol. 4, Spring, 1962. Pp. 142-145.

Borg, Walter R. *Educational Research*. New York: D. McKay Company, 1963. Pp. 338-339.

Boutwell, William D. What is meant by 'Hawthorne Effect.' *The PTA Magazine*, Vol. 58, March 1964. P. 14.

Bowers, Norman D., & Soar, Robert S. *Studies of Human Relations in the Teaching-Learning Process*. Final Report, Cooperative Research Project No. 469, U. S. Office of Education, Washington, D. C., 1961. Pp. 142-144.

Braddock, Richard, Lloyd-Jones, Richard, & Schoer, Lowell. *Research In Written Composition*, National Council of Teachers of English, Champaign, Illinois, 1963. Pp. 25-26.

Brickell, Henry M. *Organizing New York State for Educational Change*. State Department of Education, Albany, New York, 1961. Pp. 35-36.

Chall, Jeanne. *Interim Report on a Study of Methods of Teaching Beginning Reading*. Carnegie Foundation for the Advancement of Teaching, Washington, D. C., 1962.

Clark, David L. *The Hawthorne Effect and Educational Experimentation*. Address delivered to the Research Commission of the American Vocational Education Association, December 1961. (Mimeo.)

Clark, Harold. Cost and Quality in Public Education. Syracuse, New York: Syracuse University Press, 1963. P. 54.

Cook, Desmond L. The Experiment As Seen From the Subject's Viewpoint. Occasional Paper 64-158, Bureau of Educational Research and Service, The Ohio State University, 1964.

Cook, Desmond L. The Hawthorne Effect in Educational Research. Phi Delta Kappan, Vol. 44, December 1962. Pp. 116-122.

Cook, Desmond L. The Hawthorne Effect and Reading Research. In Figurel, J. Allen (Ed.), Improvement of Reading Through Classroom Practice. International Reading Association Conference Proceedings, Vol. 9, 1964. Pp. 249-253.

Cook, Desmond L. The Hawthorne Effect: Fact or Artifact, Paper presented at the 1966 Annual Meeting of the American Educational Research Association.

Cook, Desmond L. The Relationship Between Possible Hawthorne Effect Components and Stages of Experimental Investigation. Occasional Paper 63-133, Bureau of Educational Research and Service, The Ohio State University, 1963.

Coon, Lewis. The Role of Direct and Indirect Cues in a Field Experiment. Paper presented at the 1966 Annual Meeting of the American Educational Research Association.

Cooper, James G. Effects of Different Amounts of First-Grade Oral English Instruction Upon Later Reading Progress With Chamorro-Speaking Children. Journal of Educational Research, Vol. 58, No. 3, November 1964. Pp. 123-127.

Crosby, Gwladys, Fremont, Herbert & Mitzel, Harold. Mathematics Individual Learning Experiment, Final Report, Project No. 391, Title VII, NDEA Act of 1958. Pp. 22-25.

Downing, J. A. Is a Mental Age of Six Essential for Reading Readiness? Educational Research, Vol. 6, No. 1, November 1963. P. 21.

Downing, John and Halliwell, Stanley. The i.t.a. Reading Experiment in Britain. In Figurel, J. Allen (Ed.), Improvement of Reading Through Classroom Practice, International Reading Conference Proceedings, Vol. 9, 1964. P. 261.

Entwisle, Doris. Attensity: Factors of Specific Set in School Learning. Harvard Educational Review, Vol. 31, 1961. Pp. 84-101.

Entwisle, Doris. Set-Learning in Gifted High-School Students. School Review, Vol. 70. Pp. 478, 480.

Entwisle, Doris, & Forsyth, Daniel F. Word Associations of Children: The Effect of Method of Administration. Paper prepared by authors. U. S. Office of Education Grant #A7-5348.

Forehand, Charles A. Psychology of Measurement and Research on the Teaching of English. Needed Research on the Teaching of English, Cooperative Research Monograph No. 11, Office of Education, U. S. Department of Health, Education, and Welfare, Washington, D. C., 1962.

French, John. Experiments in Field Setting. In Festinger, L., & Katz, D. (Ed.), Research Methods in the Behavioral Sciences. New York: Dryden Press, 1953. Chapter 3.

Hawthorne Effect. Newsweek, April 15, 1963. P. 106.

The Hawthorne Effect in The Shape of Education for 1963-64. National School Public Relations Association, Washington, D. C., 1963.

Holt, Howard. The Educational Uses of Change. Phi Delta Kappan, 46, October 1964. Pp. 188-189.

Kaplan, Bernard A. Issues in Educating the Culturally Disadvantaged. Phi Delta Kappan, Vol. 45, November 1963. Pp. 70-76.

Keppel, Francis. Educational Statistics and Educational Policy. American Statistician, Vol. 17, December 1963. P. 11.

Kormandy, Edward J. and Van Atta, Ellis L. Experiment in Self-Instruction in General Biology. *Ohio Journal of Science*. Vol. 62, No. 1, January 1962. P. 9.

McDonald, Arthur S. Some Pitfalls in Evaluating Progress in Reading Instruction. *Phi Delta Kappan*, April 1964. Pp. 336-338.

McDonald, Arthur S. *The Placebo Response and Reading Research*. Paper presented at the Twelfth Annual Meeting of the National Reading Conference. (To be published in the *Twelfth Yearbook* of the National Reading Conference.)

McKeachie, W. J. Procedures and Techniques of Teaching: A Survey of Experimental Studies. In Sanford, H. (Ed.), *The American College*. New York: John Wiley and Sons, 1962. Chapter 8.

Miles, M. B. *Innovation in Education*. Bureau of Publications, Teachers College, Columbia University, 1964. Chapters 1, 19, Pp. 11, 461-462.

Muktarian, Herbert. Awareness of Experimental Participation and Its Relation to Achievement. Paper presented at the 1966 Annual Meeting of the American Educational Research Association.

Nelson, Kenneth G., & Pella, Milton O. Evaluation of the Adapted Harvey White Physics Films in Turkey. *Science Education*, Vol. 45, October 1961. Pp. 289-290.

Pella, Milton O. Criteria for Good Experimental Research in the Teaching of Science. *Science Education*, Vol. 45, No. 5, December 1961. Pp. 398-399.

Pella, Milton O., Stanley, Julian, Wedemeyer, C. A., & Wittich, W. A. The Use of the White Films in the Teaching of Physics. *Science Education*, Vol. 46, No. 1, February 1962. Pp. 6-21.

Porter, Douglas. *An Application of Reinforcement Principles to Classroom Teaching*. Laboratory for Research in Instruction, Harvard University, 1961. Pp. 37-38.

Psychological Research in Education, Publication 643, National Academy of Sciences, National Research Council, Washington, D. C., 1958. Pp. 19-24.

Raygor, Alton L., & Wark, David M. An Evaluation of Programmed Learning in Teaching Reading Comprehension. The National Reading Conference Yearbook, No. 11, 1961. P. 72.

Rimoldi, H. J. A., et al. Training in Problem Solving. Loyola Psychometric Laboratory, Loyola University, Publication #27, 1962.

Robinson, H. Alan. Critique of Current Research in College and Adult Reading. The National Reading Conference Yearbook, No. 13, 1963. P. 172.

Rosenbloom, Paul C. Large Scale Experimentation with Mathematics Curriculum in Research Design and Analysis, Second Annual Phi Delta Kappa Symposium on Educational Research, Phi Delta Kappa, Bloomington, Indiana, Chapter 1B, Pp. 11-46.

Schramm, Wilbur. Programmed Instruction Today and Tomorrow. Fund for the Advancement of Education, November 1962. P. 47.

Sjogren, Douglas D. The Use of Programmed Instruction in High School Correspondence Courses. Project Proposal to Cooperative Research Program, U. S. Office of Education, Washington, D. C., 1961.

Smith, Gerald R. A Picture of Research As Reflected in Proposals Submitted to the Cooperative Research Program. Mimeo. undated.

Thorndike, Robert L. The Concepts of Over- and Under-Achievement. Columbia University Press, Teachers College, New York, 1963. Pp. 31-32.

Uttal, W. R. On Conversational Interaction. In Coulson, J. E. (Ed.), Programmed Learning and Computer Based Instruction. New York, 1961. P. 190.

Veatch, Jeannette. In Defense of Individualized Reading. Elementary English. Vol. 37, 1960. P. 233.

Wallen, Norman E., & Travers, Robert M. W. Analysis and Investigation of Teaching Methods. In Gage, N. L. (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally and Company, 1963. P. 473.

Watson, Fletcher G. Research on Teaching Science. In Gage, N. L. (Ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally and Company, 1963. Chapter 20, Pp. 1047-1048.

Wilberman, Harry F. Characteristics of Some Recent Studies of Instructional Methods. In Coulson, John E. (Ed.), *Programmed Learning and Computer Based Instruction*. Proceedings of the Conference on Application of Digital Computers to Automated Instruction. New York: John Wiley and Sons, 1962.

Withall, John, & Lewis W. W. Social Interaction in the Classroom. In Gage, N. L. (Ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally and Company, 1963. Pp. 694-695.

C. References on research methodology and experimental design containing discussions having relevancy to the Hawthorne effect concept.

Aaron, J. E., Goodwin, Frances, & Kent, Vada. Fourth-Grade Teachers Experiment with Cross-Class Grouping for Reading Instruction. *Elementary English*, Vol. 36, May 1959. Pp. 305-307.

Abelson, Harold H. *The Art of Educational Research*. World Book Company, Yonkers - on - Hudson, New York, 1933. P. 144.

Behaviorial Research in Pre-clinical Psychopharmacology: Issues of Design and Technique. Public Health Service Publication No. 968, U. S. Department of Health, Education, and Welfare, Washington, D. C., 1962.

Boerk, Clarence H. The Relative Efficiency of Reading and Demonstration Methods of Instruction in Developing Scientific Understanding. *Science Education*, Vol. 40, March 1956. Pp. 92-97.

Brown, Kenneth E., Kinsella, John J., Fehr, Howard, Myers, Sheldon, & Weaver, Fred. *Analysis of Research in the Teaching of Mathematics 1959-1960*. U. S. Office of Education, Bulletin No. 12, Washington, D. C., 1963. Pp. 19, 56.

Brownell, W. A. A Critique of Research on Learning and Instruction in the School. In Henry, H. B. (Ed.), Fiftieth Yearbook of the National Society for Study of Education, Pt. 1, Graduate Study in Education. Chicago, Illinois: University of Chicago Press, 1951. Chapter 6.

Campbell, D. T. Factors Relevant to the Validity of Experiments in Social Settings. Psychological Bulletin, Vol. 54, 1957. Pp. 297-312.

Campbell, D. T., & Stanley, J. C. Experimental Designs for Research on Teaching. In Gage, N. L. (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally and Company, 1963. Chapter 5.

Cornog, June R. How People Influence Experimental Results. Paper presented at the 1963 Meeting of The American Statistical Association, Cleveland, Ohio.

Crawford, C. C. The Technique of Research in Education. University of Southern California, Los Angeles, 1928. Chapter 2.

Crespi, L. P. The Interview Effect in Polling. Public Opinion Quarterly, Vol. 12, 1948. Pp. 99-111.

Cronbach, Lee J. Course Improvement Through Evaluation. Teachers College Record, Vol. 64, May 1963. Pp. 675-677.

Durrell, Donald D. Adapting Instruction to the Learning Needs of Children in the Intermediate Grades. Journal of Education, Vol. 142, 1959. Pp. 1-78.

Eckelberry, R. H. An Uncontrolled Variable. Educational Research Bulletin, Vol. 23, January 1944. Pp. 23-24.

Engelhart, Max D. Suggestions with Respect to Experimentation Under School Conditions. Journal of Experimental Education, Vol. 14, 1946. P. 226.

Entwisle, Doris. Interaction Effects of Pre-Testing. Educational and Psychological Measurement, Vol. 21, Autumn 1961. Pp. 607-620.

Herzog, E. Some Guide Lines for Evaluative Research. Children's Bureau, Social Security Administration, U. S. Office of Health, Education, and Welfare, Washington, D. C., 1959.

Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. Experiments on Mass Communication. Princeton: Princeton University Press, 1949.

Hurst, Paul M., Jr. Relative Effectiveness of Verbal Introduction to Kinescope Recordings and Training Films. Special Devices Center, Report #269-7-42, Pennsylvania State University, May 6, 1955.

Jackson, R. Learning from Kinescopes and Films. Special Devices Center, Report #10-TV-1. Abstract in Kumata, Hideya. An Inventory of Instructional Television Research, Ann Arbor, Michigan. (no date)

Kaplan, Abraham. The Conduct of Inquiry. Chandler Publishing Co., San Francisco, 1964.

Kishka, Joseph. Describing the Effects of Experimentation in Teaching Science at the Eighth Grade Level. Science Education, Vol. 42, October 1958. P. 332.

Levitt, E. E. Clinical Research Design and Analysis. Springfield, Illinois: C. C. Thomas Company, 1961. Chapter 8.

Mason, John M., & Warrington, Willard G. An Experiment in Using Current Scientific Articles in Classroom Teaching. Science Education, Vol. 38, October 1954.

Matarazzo, J. D., Saslow, G., & Pareis, E. N. Verbal Conditioning of Two Response Classes: Some Methodological Considerations. Journal of Abnormal and Social Psychology, Vol. 61, 1960. Pp. 190-206.

McCall, W. A. How to Experiment in Education. New York: Macmillan and Company, 1923. Chapter 4.

McKeachie, W. J. Problems and Perils in Controlled Research on Teaching. Needed Research in the Teaching of English. Cooperative Research Monograph No. 11, U. S. Office of Education, Washington, D. C., 1963.

Mech, E. Victor. Performance in a Verbal Addition Task Related to Pre-experimental 'Set' and Verbal Noise. Journal of Experimental Education, Vol. 22, 1953. P. 17.

Merton, Robert K. Social Theory and Social Structure. Glencoe, Illinois: The Free Press, 1957. Pp. 421-427.

Meyers, C. E. Non-Experimental Factors and Their Control in Classroom Experimentation. Master's thesis, University of Illinois, 1939. P. 84.

Monroe, Walter S. Controlled Experimentation as a Means of Evaluating Methods of Teaching. Review of Educational Research, Vol. 4, 1934. Pp. 36-42.

Monroe, Walter S., & Engelhart, Max D. Study of the Effect of a Specified Change on a Given Cause. The Scientific Study of Educational Problems. New York: Macmillan and Company, 1936. Chapter 9, Pp. 276-277, 283-285, 298-300.

Nash, Harvey. The Design and Conduct of Experiments on the Psychological Effects of Drugs. In Uhr, L., & Miller, J. (Ed.), Drugs and Behavior. New York: John Wiley and Sons, 1960.

Orne, M. T. On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications. The American Psychologist, Vol. 17, November 1962. Pp. 776-783.

Orne, M. T. The Nature of Hypnosis: Artifact and Essence. Journal of Abnormal and Social Psychology, Vol. 58, 1949. Pp. 277-299.

Pierce, A. H. The Subconscious Again. Journal of Philosophy, Psychology, and Scientific Method, Vol. 5, 1908. Pp. 264-271.

Popham, W. James. The Influence of Novelty Effect Upon Teaching Machine Learning. Final Report, Research Grant No. 7-12-1230-155.0. Office of Education, U. S. Department of Health, Education, and Welfare, Washington, D. C., August 1962.

Reicken, Henry W. A Program for Research on Experiment in Social Psychology. In Washburne, N. F. (Ed.), Decisions, Values and Groups, New York: Pergamon Press, Inc., 1962. Pp. 102-104.

Riessmann, Frank. The Culturally Deprived Child. New York: Harper and Row, 1962. Pp. 102-104.

Robinson, Jack H. Effect of Teaching with Science Articles. Science Education, Vol. 47, February 1963. Pp. 73-83.

Rosenthal, Robert. On the Social Psychology of the Psychological Experiment: the experimenter's hypothesis as unintended determinant of experimental results. American Scientist, Vol. 51, June 1963. Pp. 268-283.

Rosenthal, Robert, & Frank, J. O. Psychotherapy and the Placebo Effect. Psychological Bulletin, Vol. 53, 1956. Pp. 294-302.

Rosenthal, Robert. Note on the Fallible E. Psychological Reports, Vol. 4, 1958. P. 662.

Ross, Sherman et al. Drugs and Placebos: A Model Design. Psychological Reports, Vol. 10, 1962. Pp. 383-392.

Smith, B. O. Science of Education in Encyclopedia of Educational Research. New York: Macmillan and Company, 1952. Pp. 148-149.

Star, S. A., & Hughes, H. M. Report on an Educational Campaign: The Cincinnati Plan for the United Nations. American Journal of Sociology, Vol. 55, 1949-50. P. 389.

Strang, Ruth. Reactions to Research on Reading. Educational Forum, Vol. 26, January 1962. Pp. 187-192.

Talmadge, Max, Davids, Anthony, & Laufer, M. W. A Study of Experimental Methods for Teaching Emotionally Disturbed, Brain Damaged, Retarded Readers. The Journal of Educational Research, Vol. 56, 1963. P. 315.

Travers, Robert M. W. An Introduction to Educational Research. New York: Macmillan and Company, 1958. Pp. 336-363.

Travers, Robert M. W. A Study of the Relationship of Psychological Research to Educational Practice. In Glaser, Robert (Ed.), Training Research and Education. University of Pittsburgh Press, 1962.

Washington County, Maryland Board of Education. Washington County Closed Circuit Television Report, Washington County, Maryland, (no date).

II. Experimenter Bias

Brogden, W. J. The Experimenter as a Factor in Animal Conditioning. Psychological Reports, Vol. 2, 1962 Pp. 239-242.

Dohrenwend, B. S., & Richardson, S. A. Directiveness and Non-directiveness in Research Interview: A Reformulation of the Problem. Psychological Bulletin, Vol. 60, No. 5, September 1963. Pp. 475-485.

Ekren, Ursula. The Effect of Experimenter Knowledge of a Subject's Scholastic Standing on the Performance of a Reasoning Task. Unpublished Master's thesis, Marquette University, 1962.

Ferber, R., & Wales, H. G. Detection and Correction of Interviewer Bias. Public Opinion Quarterly, Vol. 16, 1952. Pp. 107-127.

Fode, K. L. The Effect of Experimenter's and Subject's Level of Anxiety and Need for Approval on Experimenter Outcome-Bias. Unpublished doctoral dissertation, University of North Dakota, 1963.

Fode, K. L. The Effect of Non-Visual and Non-Verbal Interaction on Experimenter Bias. Unpublished Master's thesis, University of North Dakota, 1960.

Freiberg, A. D., Vaughn, C. L., & Evans, M. C. Effect of Interviewer Bias upon Questionnaire Results Obtained with a Large Number of Investigators. American Psychologist, Vol. 1, 1946. P. 243.

Friedman, P. A Second Experiment on Interviewer Bias. Sociometry, Vol. 5, 1942. Pp. 378-379.

Hildum, D. C., & Brown, R. W. Verbal Reinforcement and Interviewer Bias. Journal of Abnormal and Social Psychology, Vol. 53, 1956. Pp. 108-111.

Krasner, Leonard. A Technique for Investigating the Relationship Between the Behavior Cues of the Examiner and the Verbal Behavior of the Patient. *Journal of Consulting Psychology*, Vol. 22, No. 5, 1958. Pp. 264-366.

Krasner, Leonard, Ullman, L., Weiss, R. L. & Collins, B. J. Responsivity to Verbal Conditioning as a Function of Three Different Examiners. *Journal of Clinical Psychology*, Vol. 17, No. 4, October 1961. Pp. 411-415.

Lindzey, G. A Note on Interviewer Bias. *Journal of Applied Psychology*, Vol. 35, 1951. Pp. 182-184.

Marcia, J. *Hypothesis-making, Need for Social Approval, and Their Effects on Conscious Experimenter Bias*. Unpublished Master's thesis, The Ohio State University, 1961.

McGuigan, F. M. The Experimenter: A Neglected Stimulus Object. *Psychological Bulletin*, Vol. 60, No. 4, July 1963. Pp. 421-428.

McLean, Leslie D. *The Problems of Experimenter Bias*. Occasional Paper 64-158, Bureau of Educational Research and Service, The Ohio State University, 1964.

Mills, T. A Sleeper Variable in Small Groups Research: The Experimenter. *Pacific Sociological Review*, Vol. 5, 1962. Pp. 21-28.

Mosteller, F. Correcting for Interviewer Bias. In Cantril, H. (Ed.), *Gauging Public Opinion*. Princeton University Press, 1944. Pp. 286-288.

Pflugrath, J. *Examiner Influence in a Group Testing Situation with Particular Reference to Examiner Bias*. Unpublished Master's thesis, University of North Dakota, 1962.

Rosenthal, Robert. Experimenter Attributes as Determinants of Subjects' Responses. *Journal of Projective Techniques and Personality Assessment*, Vol. 27, No. 3, 1963. Pp. 324-331.

Rosenthal, Robert. Experimenter Modeling Effects as Determinants of Subjects' Responses. *Journal of Projective Techniques and Personality Assessment*, Vol. 27, No. 4, 1963. Pp. 467-471.

Rosenthal, Robert. Experimenter Outcome-Orientation and the Results of the Psychological Experiment. Psychological Bulletin, Vol. 61, No. 6, June 1964. Pp. 405-411.

Rosenthal, Robert. On the Social Psychology of the Psychological Experiment: With Particular Reference to Experimenter Bias. Paper presented at the American Psychological Association annual convention, New York, N. Y., September 1961.

Rosenthal, Robert. Projection, Excitement and Unconscious Experimenter Bias. The American Psychologist, Vol. 13, 1958. Pp. 345-346, (Abstract) (b)

Rosenthal, Robert. The Effect of Experimenter Outcome-Bias and Subject Set on Awareness in Verbal Conditioning Experiments. Journal of Verbal Learning and Verbal Behavior, Vol. 2, No. 3, October 1963. Pp. 275-283.

Rosenthal, Robert. The Hypothesis of Unconscious Experimenter Bias. In: An Attempt at the Experimental Induction of the Defense Mechanism of Projection. Unpublished doctoral dissertation, University of California, Los Angeles, 1956.

Rosenthal, Robert, & Fode, K. L. Psychology of the Scientists: V. Three Experiments in Experimenter Bias. Psychological Reports, Vol. 12, No. 3, 1963. Pp. 491-511.

Rosenthal, Robert, & Fode, K. L. The Effect of Experimenter Bias on the Performance of the Albino Rat. Behavioral Science, Vol. 8, No. 3, July 1963. Pp. 183-189.

Rosenthal, Robert, & Fode, K. L. The Problem of Experimenter Outcome-Bias. In Ray, D. P. (Ed.), Series Research in Social Psychology. Symposia studies series No. 8, National Institute of Behavioral Sciences, Washington, D. C., 1961.

Rosenthal, Robert, Fode, K. L., Friedman, C. J., & Vikan-Kline, Linda L. Subject's Perception of their Experimenter under Conditions of Experimenter Bias. Perceptual Motor Skills, Vol. 11, 1960. Pp. 325-331.

Rosenthal, Robert, Fode, K. L., & Vikan-Kline, Linda L. *The Effect on Experimenter Bias of Varying Levels of Motivation of Es and Ss*. Unpublished manuscript, Harvard University, 1960.

Rosenthal, Robert, Fode, K. L., Vikan-Kline, Linda L., & Persinger, G. W. Verbal Conditioning: Mediator of Experimenter-Expectancy Effects. *Psychological Reports*, Vol. 14, 1964. Pp. 71-74.

Rosenthal, Robert, Friedman, C. J., Johnson, C. A., Fode, K. L., Schill, T. R., White, C. R., & Vikan-Kline, Linda L. Variables Affecting Experimenter Bias in a Group Situation. *Genetic Psychology Monogram*, in press.

Rosenthal, Robert, & Gaito, John. The Interpretation of Levels of Significance by Psychological Researchers. *Journal of Psychology*, Vol. 55, 1963. Pp. 33-38.

Rosenthal, Robert, & Halas, E. S. Experimenter Effect in the Study of Invertebrate Behavior. *Psychological Reports*, Vol. 11, 1962. Pp. 251-256.

Rosenthal, Robert, & Lawson, R. *A Longitudinal Study of the Effects of Experimenter Bias on the Operant Learning of Laboratory Rats*. Unpublished manuscript, Harvard University, 1961.

Rosenthal, Robert, & Persinger, G. W. Let's Pretend: Subject's Perception of Imaginary Experimenters. *Perceptual Motor Skills*, Vol. 14, 1962. Pp. 407-409.

Rosenthal, Robert, Persinger, G. W., & Fode, K. L. Experimenter Bias, Anxiety, and Social Desirability. *Perceptual Motor Skills*, Vol. 15, 1962. Pp. 73-74.

Rosenthal, Robert, Persinger, G. W., Vikan-Kline, Linda L., & Fode, K. L. The Effect of Early Data Returns on Data Subsequently Obtained by Outcome-Biased Experimenters. *Sociometry*, Vol. 26, No. 4, December 1963.

Rosenthal, Robert, Persinger, G. W., Vikan-Kline, Linda L., & Fode, K. L. The Effect of Experimenter Outcome-Bias and Subject Set on Awareness in Verbal Conditioning Experiments. *Journal of Verbal Learning and Verbal Behavior*, Vol. 2, No. 3, October 1963.

Rosenthal, Robert, Persinger, G. W., Vikan-Kline, Linda L., & Mulry, R. C. The Role of the Research Assistant in the Mediation of Experimenter Bias. Journal of Personality, Vol. 31, No. 3, September 1963.

Smith, H. L., & Hyman, H. The Biasing Effect of Interviewer Expectations on Survey Results. Public Opinion Quarterly, Vol. 14, 1950. Pp. 491-506.

Stanton, F., & Baker, K. H. Interviewer Bias and the Recall of Incompletely Learned Materials. Sociometry, Vol. 5, 1942. Pp. 123-234.

Ullman, L. P., Krasner, L., & Collins, B. J. Modifications of Behavior through Verbal Conditioning: Effects in Group Therapy Journal of Abnormal and Social Psychology, Vol. 62, No. 1, 1961. Pp. 128-132.

Vikan-Kline, Linda L. The Effect of an Experimenter's Perceived Status on the Mediation of Experimenter Bias. Unpublished Master's thesis, University of North Dakota, 1962.

Weitman, Morris. An Examination of Some Variables Presumably Relevant to Induced Change in the Interview. Abstract of paper presented at Western Psychological Association meeting, Santa Monica, April 19, 1963.

White, C. R. The Effect of Induced Subject Expectations on the Experimenter Bias Situations. Unpublished doctoral dissertation, University of North Dakota, 1962.

Wickes, T. A., Jr. Examiner Influence in a Testing Situation. Journal of Consulting Psychology, Vol. 20, 1956. Pp. 23-26.

III. Awareness

Adams, J. Laboratory Studies of Behavior Without Awareness. Psychological Bulletin, Vol. 54, 1957. Pp. 151-171.

Adrion, Jo Ann. The Problem of Awareness and Its Relationship to Counseling, Unpublished Master's thesis, The Ohio State University, 1964.

Arginteanu, J. Verbal Conditioning, Awareness, and Their Relation as a Function of Measures of Social Desirability and Resistance to Verbal Conditioning. Unpublished doctoral dissertation, State University of Iowa, 1961.

Assenrath, Julius M. Transfer of Learning Without Awareness. *Psychological Reports*, Vol. 10, 1962. Pp. 411-429.

Beatty, Walcott. Is There Verbal Conditioning Without Awareness? San Francisco State College, San Francisco, California. (Mimeo.)

Brown, F. J. Knowledge of Results as an Incentive in Schoolroom Practice. *Journal of Educational Psychology*, Vol. 23, October 1952. Pp. 532-552.

Coon, Lewis H. *Problems of Measuring Awareness of Experimentation*. Occasional Paper 64-158, Bureau of Educational Research and Service, The Ohio State University, 1964.

Daily, J. M. Verbal Conditioning Without Awareness. *Dissertation Abstracts*, Vol. 12, 1953. Pp. 1247-1248.

Dulaney, D. D. The Place of Hypotheses and Intentions: An Analysis of Verbal Control in Verbal Conditioning. In Eriksen, C. W. (Ed.), *Behavior and Awareness - a symposium of research and interpretation*. Durham, N. C.: Duke University Press, 1962. Pp. 102-129.

Ekman, P. *A Comparison of Verbal and Non-Verbal Behavior as Reinforcing Stimuli of Opinion Responses*. Unpublished doctoral dissertation, Adelphi College, Garden City, New York, 1958.

Ekman, Paul, Krasner, Leonard, & Ullman, Leonard. Interaction of Sex and Awareness as Determinant of Response to Verbal Conditioning. *Journal of Abnormal and Social Psychology*, 1963. Pp. 387-389.

Eriksen, C. W. Discrimination and Learning Without Awareness: A Methodological Survey and Evaluation. *Psychological Review*, Vol. 67, 1960. Pp. 279-300.

Eriksen, C. W. Figments, Fantasies and Follies: A Search for the Subconscious Mind. Behavior and Awareness - a symposium of research and interpretation. Durham, N. C.: Duke University Press, 1962.

Eriksen, C. W., & Kuethe, J. L. Avoidance Conditioning of Verbal Behavior Without Awareness: A Paradigm of Repression. Journal of Abnormal Psychology, Vol. 53, 1956. Pp. 203-209.

Essman, W. B. Awareness of Reinforcement in Learning Without Awareness. Psychological Reports, Vol. 3, 1957. Pp. 399-400.

Farber, I. E. The Things People Say to Themselves. The American Psychologist, Vol. 18, No. 4, April 1963. Pp. 185-197.

Frank, Jerome D. Experimental Studies of Personal Pressure and Resistance. I. Experimental Production of Resistance. Journal of General Psychology, 1944. Pp. 23-41.

French, J. R. P. Group Dynamics, Draft of a Report for a Course on Methodology, University of Michigan, Ann Arbor, Spring 1963.

Gardner, E. F., & Thompson, George G. Social Relations and Morale in Small Groups. New York: Appleton-Century-Crofts, 1956.

Gerstein, A. I. The Effect of Reinforcement Schedule on Meaning Generalization and on Awareness of the Purpose of the Experiment. Journal of Personality, Vol. 29, September 1961. Pp. 350-362.

Greenspoon, J. The Reinforcing Effect of Two Spoken Sounds on the Frequency of Two Responses. American Journal of Psychology, Vol. 68, 1955. Pp. 409-416.

Hilgard, Ernest R. What Becomes of the Input from the Stimulus. In Eriksen, C. W. (Ed.), Behavior and Awareness - a symposium of research and interpretation. Durham, N. C.: Duke University Press, 1962.

Hills, Barbara B. A Study of Acquisition, Extinction, and Awareness in Verbal Conditioning. Unpublished Master's dissertation, State University of Iowa, 1959.

Hirsch, J. Learning Without Awareness and Extinction Following Awareness as a Function of Reinforcement. *Journal of Experimental Psychology*, Vol. 54, 1957. Pp. 218-224.

Irwin, F. W., Kauffman, K., Prior, G., & Weaver, H. B. On Learning Without Awareness of What is Being Learned. *Journal of Experimental Psychology*, Vol. 17, 1934. Pp. 823-827.

Kelley, H. H. The Effects of Expectations upon Impressions of Persons. *The American Psychologist*, Vol. 4, 1949. P. 252.

Kimble, G. A. Classical Conditioning and the Problem of Awareness. In Eriksen, C. W. (Ed.), *Behavior and Awareness - a symposium of research and interpretation*. Durham, N. C.: Duke University Press, 1962. Pp. 27-45.

Kranfer, F. H., & Marston, A. R. Verbal Conditioning, Ambiguity, and Psychological Therapy. *Psychological Reports*, Vol. 9, 1961. Pp. 461-475.

Krasner, L. A Technique of Investigating the Relationships Between Behavior Cues of Examiner and Verbal Behavior of Patient. *Journal of Consulting Psychology*, Vol. 22, No. 5, 1958. Pp. 364-366.

Krasner, L. Studies of the Conditioning of Verbal Behavior. *Psychological Bulletin*, Vol. 55, 1958. Pp. 148-170.

Krasner, L. *The Behavioral Scientist and Social Responsibility: No Place To Hide*. Paper presented at The American Psychological Association Annual Meeting, Philadelphia, September 1963.

Krasner, L., & Ullman, Leonard. Variables Affecting Report of Awareness in Verbal Conditioning. *Journal of Psychology*, Vol. 56, 1963. Pp. 199-202.

Krasner, L., Ullman, Leonard P., Weiss, Robert L., & Collins, Beverly J. Responsivity to Verbal Conditioning As a Function of Three Different Examiners. *Journal of Clinical Psychology*, Vol. 17, No. 4, October 1961. Pp. 411-415.

Krasner, L., Weiss, Robert L., & Ullman, Leonard P. Responsivity to Verbal Conditioning as a Function of Awareness. Psychological Reports, Vol. 8, 1961. Pp. 523-538.

Krieckhaus, E. E., & Eriksen, C. W. A Study of Awareness and Its Effects on Learning and Generalization. Journal of Personality, Vol. 28, 1960. Pp. 503-517.

Lana, Robert E., & Rosnow, R. L. Subject Awareness and Order Effects in Persuasive Communications. Psychological Reports, Vol. 12, 1963. Pp. 523-529.

Lanyon, R. I. Verbal Conditioning and Awareness in a New Sentence Construction Task. Unpublished Master's dissertation, State University of Iowa, 1962.

Lawrence, Frederic K. Measuring Subject Awareness of Experimentation. Unpublished Master's thesis, The Ohio State University, 1965.

Levin, Saul M. The Effects of Awareness on Verbal Conditioning. Journal of Experimental Psychology, Vol. 61, No. 1, 1961. Pp. 67-75.

Matarazzo, J. D., Weitman, M., Saslow, G., & Wiens, A. N. Interviewer Influence on Durations of Interviewee Speech. Journal of Verbal Learning and Verbal Behavior, Vol. 1, No. 6, May 1963. Pp. 451-458.

McKinley, R. A. Awareness, Personality Variables, and Performance in Verbal Conditioning. Unpublished Master's dissertation, State University of Iowa, 1962.

Mechanic, A. An Investigation of the Sets Aroused by Intentional Learning Instructions. Paper presented at the Western Psychological Association, Santa Monica, April 1963.

Mechanic, A. The Responses Involved in the Rote Learning of Verbal Materials. Journal of Verbal Learning and Verbal Behavior, Vol. 3, February 1964. Pp. 30-36.

Nielson, A. M. Awareness of In-group Attitudes. Doctoral dissertation, The Ohio State University, 1955. 110 pages.

O'Hanlon, J. Jr., Schmidt, E. A., & Baker, C. H. A Study to Determine the Effects of Placebos Upon Performance in a Vigilance Task. Human Factors Research, Inc. Report 750-3, January 1964.

Palermo, D. S., & Jenkins, J. Word Association Norms: Grade School through College. Minneapolis: University of Minnesota Press, 1964.

Panlasigui, I., & Knight, F. B. The Effect of Awareness of Success or Failure. National Society for the Study of Education, 29th Yearbook, Bloomington, Illinois: Public School Publishing Company, 1930. Pp. 611-620. Chapter 11.

Philbrick, E. B., & Postman, L. A Further Analysis of Learning Without Awareness. American Journal of Psychology, Vol. 68, 1955. Pp. 417-424.

Phillips, J. S., Matarazzo, J. D., Matarazzo, R. G., & Saslow, G. Observer Reliability of Interaction Patterns during Interviews. Journal of Consulting Psychology, Vol. 21, 1957. Pp. 269-275.

Postman, L., & Jarrett, R. F. An Experimental Analysis of Learning Without Awareness. American Journal of Psychology, Vol. 65, 1952. Pp. 244-255.

Ronning, R. R. Application of Multi-dimensional and Scale Analysis to Interest Measurement, Cooperative Research Project, University of California, Berkely, 1963.

Ross, C. C. The Influence of Achievement Upon Knowledge of Progress. Journal of Educational Psychology, Vol. 24, November 1933. Pp. 609-611.

Sidowski, J. B. Influence of Awareness of Reinforcement on Verbal Conditioning. Journal of Experimental Psychology, Vol. 48, 1954. Pp. 355-360.

Southwell, Eugene. A Note on Awareness in Verbal Conditioning. Psychological Reports, Vol. 10, 1962. P. 40.

Spielberger, C. D. The Role of Awareness in Verbal Conditioning. In Eriksen, C. W. (Ed.), Behavior and Awareness - a symposium of research and interpretation. Durham, N. C.: Duke University Press, 1962. Pp. 73-101.

Spielberger, C. D., Levin, S. M., & Shepard, M. The Effects of Awareness and Attitude Toward the Reinforcement on the Operant Conditioning of Verbal Behavior. *Journal of Personality*, Vol. 30, 1962. Pp. 106-121.

Spires, A. M. *Subject-Experimenter Interaction in Verbal Conditioning*. University Microfilms, Inc., Ann Arbor, Michigan 1960.

Tatz, S. J. *Symbolic Mediation in Learning Without Awareness*. Paper presented at the Eastern Psychological Association Meeting, Atlantic City, March 1956.

Thorndike, E. L., & Rock, R. T., Jr. Learning Without Awareness of What is Being Learned or Intent to Learn It. *Journal of Experimental Psychology*, Vol. 17, 1954. Pp. 1-19.

Ullman, Leonard, & Krasner, L. The Placebo Response: An Experimental Approach. *Journal of Nervous and Mental Diseases*, Vol. 136, 1964. Pp. 379-387.

Ullman, Leonard, & Krasner L., & Collins, Beverly J. Modifications of Behavior Through Verbal Conditioning: Effects in Group Therapy. *Journal of Abnormal and Social Psychology*, Vol. 62, 1961. Pp. 128-132.

Ullman, Leonard, Weiss, Robert L., & Krasner, L. The Effect of Verbal Conditioning of Emotional Words on Recognition of Threatening Stimuli. *Journal of Clinical Psychology*, Vol. 19, 1963. Pp. 182-183.

Verplanck, W. S. Unaware of Where's Awareness: Some Verbal Operants - Notates, Monents, and Notants. In Eriksen, C. W. (Ed.), *Behavior and Awareness - a symposium of research and interpretation*. Durham, N. C.: Duke University Press, 1962. Pp. 130-158.

Weiss, R. L. The Influence of Set for Speed on Learning Without Awareness. *American Journal of Psychology*, Vol. 68, 1955. Pp. 425-431.

Weiss, R. L., Krasner, L., & Ullman, Leonard. Responsivity of Psychiatric Patients to Verbal Conditioning: Success and Failure Conditions and Pattern of Reinforced Trials. *Psychological Reports*, 1963. Pp. 423-426.

Appendix A

INSTRUCTIONS FOR TEACHERS OF ARITHMETIC CLASSES WHICH ARE INFORMED AS TO THEIR INVOLVEMENT IN AN EXPERIMENT

September 2, 1964

In order to best serve the purpose of the Hawthorne effect field research experiment, it is necessary to describe to you, the cooperating teacher, specific procedure which you will be asked to follow as you teach arithmetic throughout the coming school year.

The fundamental task for the teacher of an "informed" arithmetic class is to inform the students of their participation in an experiment. In order to gain initially the students' interest and cooperation in the experiment, it will be necessary for you to inform the students during arithmetic class on the first day of the experiment (Sept. 9, 1964) as to their participation in the experiment. No less important than this initial procedure, however, is the need to periodically remind your class of their participation in the experiment in order to maintain their initial interest and cooperation. Consequently, as the teacher of an "informed" arithmetic class, you will be asked, first, to inform your class at the beginning of the school year, and second, to remind them periodically throughout the school year of their participation in an experiment.

The major portion of these instructions is concerned with the methods by which the students are informed, the specific content of this information, and the schedule which is to be followed for initially informing and periodically reminding your students of their participation in an experiment.

In all probability, once you have informed your arithmetic class about the experiment and aroused their interest, you will be confronted by individual students seeking more information about the experiment than you have already given them. Additional instructions as to the procedure to follow in handling these situations will also be described in the later section.

I. Specific Content of Information Presented to "Informed" Arithmetic Classes

Because of the need for a specific content with regard to the nature of the information to be given to each "informed" arithmetic class in the experiment, the essential points which are to be stressed in the initial information and the Periodic Reminders are summarized below.

Following each list of essential points will be found suggested presentation forms which you may use as guides for presenting the specific essential points (further information on the method for presenting these essential points follow in the next section).

A. Initial Information Content

1. Essential Points:
    - a. The arithmetic period will be involved in an experiment.
    - b. By class discussion, if possible, define and describe the nature of experiment.
    - c. Teacher should summarize definition of "experiment" and give examples if possible.
    - d. Make transition between discussion of experiments in general to explanation of this particular experiment.
    - e. This experiment is set up to evaluate a different way of learning arithmetic.
    - f. Allay possible student anxiety that arithmetic will be more difficult because of participation in the experiment.
    - g. Express the need for each student's cooperation.
    - h. Explain that Periodic Reminders will be given to remind students of experimental participation.

2. Suggested Presentation Form:

In our arithmetic period (class) this year we will be taking part in an experiment. Who can tell us what an experiment is?

(At this point lead class discussion toward a definition and description of "experiment." Summarize class responses, if any, with this suggested definition: "An experiment is an attempt by a scientist to prove or disprove something which is unknown or uncertain at the present time." Expand on this definition and/or give examples at your own discretion.)

This year we are going to participate in an experiment like the one (those) we have been talking about.

The experiment of which we are a part is set up to discover the value of a different way of learning arithmetic. (It is much like the experiment your arithmetic class participated in last year.) Because we are taking part in an experiment, our learning arithmetic should be no harder than it has been for you in the past.

It is hoped that you will all cooperate with the experiment and learn as much as you can about arithmetic this year. Because some of you may sometimes forget that we are involved in an experiment, I will remind you now and then.

B. Periodic Reminder Content

1. Essential Points:
    - a. Remind students of participation in an experiment.
    - b. Experiment is set up to evaluate a different way of learning arithmetic.
    - c. Allay possible anxiety concerning increased difficulty in arithmetic because of experimental participation.
    - d. Re-emphasize need for cooperation.

2. Suggested Presentation Form:

In case some of you may have forgotten, I would like to remind all of you (again) that we are taking part in an experiment. Remember that the purpose of the experiment is to discover the value of a different way of learning arithmetic.

Remember, too, that our learning arithmetic should be no harder than it would be if we were not taking part in an experiment, and that your cooperation is needed to discover the value of the new way of learning arithmetic.

II. Method of Presenting Information to the "Informed" Arithmetic Class

The nature of this experiment and the requirement that each of the essential points listed in the previous section must be included in the information given to your class necessarily limits the methods of presentation you may use.

The one presentation method which we feel would give you the most latitude in personal expression and most natural delivery, yet would ensure the inclusion of all essential points in your presentation, is to paraphrase the presentation form examples described in the preceding section. It is essential to the success of this aspect of the experiment that, in paraphrasing the preceding example forms, you maintain the order of essential points as they are listed.

III. Schedules for Presenting Information to the "Informed" Arithmetic Class

A. Initial Information Presentation Schedule

The Initial Information concerning the arithmetic period's involvement in an experiment will be presented at the beginning of arithmetic period on the first day in the second year of the Hawthorne effect field research experiment, September 9, 1964.

B. Periodic Reminder Presentation Schedule

The Periodic Reminders will be scheduled for presentation during arithmetic period on a basis of once during the third week and once during the sixth week of every grading period during the school year. The choice of the specific day of the week on which the Periodic Reminder will be presented is at your own discretion.

In the event that an arithmetic test (a "test" in this sense meaning one which covers a chapter or more of work) is planned during the third or sixth week of a grading period, you should present the scheduled Periodic Reminder during arithmetic class on the day preceding the day of the test.

At the beginning of arithmetic period on the first day of school after Thanksgiving, Christmas, and Easter vacations, a Periodic Reminder should be presented. In the event that a "regular" Periodic Reminder is scheduled for presentation during the week immediately following the end of any of these vacation periods, this "special" reminder presentation is to be ignored. (A check of Appendix A, the calendar schedule for presentation of the Initial Information and Periodic Reminders, will allow you to forsee the possible use of "special" reminders.)

IV. General Procedure to Follow in the Event of Individual Student Questions Concerning the Experiment

In the event that you are questioned by your students as to aspects of the experiment more specific than those which have previously been mentioned in the Initial Information and Periodic Reminders, your answer to whatever question may be asked must be within the limits of the information to which your students have already been exposed. In other words, any responses you may give to students' questions concerning the experiment should not involve any information which could possibly affect adversely the results of the Hawthorne effect field research experiment.

Following are examples of questions which, because of students' natural interest and curiosity, we feel you may be confronted with. In all probability there will be many other questions asked of you concerning the experiment, but the answers which are given here we hope will serve as guidelines to orient your responses to additional questions which arise.

1. In the particular schools where two teachers exchange classrooms to teach arithmetic, the question may arise, directed to the "different" teacher, "Why are you teaching us arithmetic instead of our regular teacher, Mrs. X?" A similar question may be asked of the classes' regular teacher, as "Why is Miss Y teaching us arithmetic instead of you?"

   In both situations, the teacher's answer should inform the student that the teachers exchanging classrooms is simply a part of the experiment.

2. In those classrooms which will be using the SMSG arithmetic books, the question may arise, "Why are we using a new arithmetic book?"

   In reply to this type of question, the teacher should inform the student that the new textbook is the means by which the different way of learning arithmetic will be accomplished.

3. Some students may be curious as to, "What other schools are participating in the experiment?"

Your answer is simply, "I don't know." Do not confirm or deny the participation of schools suggested by the students.

V. Experimental Log

From time to time during the school year a student may say something which could contribute to the experiment in a way that standardized tests do not measure.

i.e. "My cousin at Clinton School is using the same 'new' textbook we are."

Please keep a record of such events in writing and pass them along to the Associate in Mathematics during his regular visits.

1964-65 CALENDAR SCHEDULE FOR PRESENTATION OF THE INITIAL INFORMATION AND PERIODIC REMINDERS

1. Date for Presentation of Initial Information:

   September 9, 1964

2. Dates for Presentation of Periodic Reminders:

   a. Week of September 28, 1964
   b. Week of October 19, 1964
   c. Week of November 9, 1964
   d. Week of November 30, 1964
   e. Week of December 20, 1964
   f. Week of January 4, 1965
   g. Week of January 25, 1965
   h. Week of February 15, 1965
   i. Week of March 8, 1965
   j. Week of March 29, 1965
   k. Week of April 19, 1965

3. Dates for Presentation of "Special" Periodic Reminders:

   This year they coincide with a schedule for cueing every three weeks.

INSTRUCTIONS TO TEACHERS OF ARITHMETIC CLASSES NOT INFORMED AS TO PARTICIPATION IN AN EXPERIMENT

September 2, 1964

Because of the remote possibility (we hope) that one or more of your students may in some manner suspect that they are involved in an experiment, it is necessary that you be forewarned and advised as to the response which you should give to questions concerning participation in an experiment.

If any student questions you, directly or indirectly, as to the involvement of your arithmetic class in an experiment, your answer is to be an unequivocal NO.

Any answer other than firmly negative may in some manner adversely affect the results of the Hawthorne effect field research experiment.

In the event an inquiry is made concerning the visits of the Associate in Mathematics, your response should be in line with the concept of his visiting your classroom regularly in order to give him a better idea as to how elementary school children learn arithmetic. His role in the classroom from the student's point of view should be much the same as a student teacher present for observation only.

Appendix B

Development of the Awareness Inventory[1]

One of the major objectives of the investigation was to ascertain the nature of a possible relationship existing between awareness of participation in an experimental situation and subsequent gains in arithmetic achievement. The substantiative hypothesis to be tested was that there would be a positive relationship between such gains and awareness of experimental participation. To test this hypothesis, it was necessary to secure from each subject some assessment of awareness of participation in an experiment. The purpose of this appendix is to describe the procedure employed to develop the instrumentation which provided a quantified statement of the knowledge of participation.

Definition of Awareness. The first step was to establish an operational definition of "awareness of experimental participation" consistent with the concept presented in Chapter I. For purposes of this investigation, such awareness was defined as the student's self-disclosure that he had been involved in an experimental situation. By self-disclosure, it was meant that the subject was not directly informed by some person that he was, or had been, participating in an experiment but rather that the student came to this conclusion by means of his own discovery. In this study, a further qualification was needed in view of the nature of the direct and indirect cues described in Chapter II. The direct cue of telling the subject would obviously be responded to positively by those receiving such cues but would not indicate the extent to which the student perceived his participation based upon other indirect cues presented to him. This latter aspect of awareness is considered by many persons to be an essential factor of the presence or absence of the Hawthorne effect in educational research projects. Consequently, it was decided to focus upon the measurement of the student's self-disclosure as a function of the two indirect cues used as independent variables in the experimental design. The first variable was the use of a new or different text for arithmetic. The second was the presence or absence of the regular classroom teacher for arithmetic instruction.

---

[1] The development of the oral structive approach was undertaken largely by Jo Ann Adrion and Fred Lawrence while the final instrumentation was developed by Herbert Muktarian. The contribution of these persons is gratefully acknowledged by the principal investigator.

Pilot Studies

A survey of the literature reviewed briefly in Chapter I on awareness in experimental investigations revealed that the oral interview of subjects subsequent to the experimental situations was the most common technique employed to assess awareness. Although no firm conclusion exists based upon the studies reviewed as to the value of the oral interview approach to assess awareness, a decision was made to pursue the use of such a technique as a means of assessing the subjects awareness of participation in the experimental setting. The problem was made still more difficult due to the lack of any published research that subjects on the age-grade level used in this study could adequately verbalize their awareness even if it did exist. The several studies reviewed in Chapter I were conducted generally on subjects of college age level and hence are of questionable value in predicting whether or not elementary school studies could make self-reports on experimental participation.

Having decided to use the oral interview technique, it was felt opportune to conduct a series of pilot studies on students comparable to those used in the experimental situation to determine (a) the adequacy of verbalization of fifth graders in an interview situation, (b) their conceptualization of an experiment, (c) the number and type of question needed in order to permit the subject to disclose he was in an experiment, and (d) if the subject's responses could be scaled to indicate degree of experimental awareness.

Verbalization. The pilot study on verbalization was undertaken at small (enrollment 540 students) middle class elementary school within the Columbus public school system. This school had in operation several teacher-planned innovations but no formal methods experiments were in process at the time of the study. A list of questions was prepared for this study as given below.

Question Set I - Sampling for Verbal Ability Level of Fifth Grade Pupils.

1. At what grade did you start at ___________?
2. What was your favorite grade? Why?
3. What are your two most favorite subjects in school?
4. (a) Do you like arithmetic? Why?

   or

   (b) Why is arithmetic one of your favorites?
5. What is an experiment? i.e. What does the word experiment mean to you?

6. What kind of people do experiments?
7. Why do they experiment?
8. Have you ever been part of an experiment? Tell me about it.
9. Do you think that your arithmetic class is part of an experiment? Why? or Why not?

The subjects were six fifth grade students and consisted of a boy and girl from each third of the IQ range of all fifth grade students in order to look for possible verbalization differences by sex and mental ability. The interviews were conducted individually with the responses tape recorded. The subjects were told that the purpose of the interview was to find out what they thought about the school.

A subjective analysis of the transcribed tapes indicated that the verbal ability of the students was sufficient to make an interview technique feasible. The vocabulary seemed extensive enough to express thoughts although at times the logical sequence of statements was transposed. It was also found with this particular group of subjects that it was not possible to elicit responses that they were participants in an experiment or if they were not. Two general observations were (1) that the presence of the tape recorder did not seem to be perceived by the students as unusual procedure since all had used one during reading class as a teaching aid for self-evaluation, and (2) the student's first response to the word "experiment" unanimously involved either a physical science or medical interpretation. On the basis of these results it was recommended that the interviews be expanded to include alternate questions for clarification and additional questions should be devised to provide a greater volume of responses for more reliable evaluation of student responses.

Number and Type of Questions. The second pilot study was conducted to derive additional guidelines with regard to developing the awareness instrument. First, as a preliminary measure prefacing the validation of the interview questions. Hopefully, this group of interviews would produce gross indications as to the appropriateness of the questions for an actual experimental group. Second, to further increase the sample of responses indicating verbal ability level of upper-grade elementary school children. Third, introducing the variable of experimentation it would be possible to allow for a comparison of responses to the interview between an experimental group and control group.

The setting for this pilot study was a small, middle-class elementary school in a suburban school system outside of Columbus. This school had in operation a highly publicized, administration planned project in team teaching. The project involved the entire fifth grade and was a modification of the Trump plan. Major variables employed were large and small group instruction, changing rooms, and teachers for different subjects, grouping for individual differences, and team-student and team-parent conferences on a regular schedule. The project had received much publicity in the community by television appearances of the teaching team, newsletters to parents, and special PTA meetings concerned only with the team-teaching project.

The subjects were selected as before from all fifth graders involved in the project. The interviews were conducted individually with responses tape recorded. The subjects were told the purpose of the interview was to find out what they thought of their school. The set of specific questions developed for this study were as follows:

Question Set II - Sampling Verbal Ability of Fifth Grade Pupils Participating in Team-teaching Project.

1. (a) What was your favorite grade?
   (b) Why?
   (c) How long have you attended this school?

2. (a) What two subjects are your favorites?
   (b) Why?

3. (a) Do you like arithmetic?
   (b) Why?

4. (a) What is an experiment?
   (b) What does the word experiment mean to you?

5. (a) What kinds of people do experiments?
   (b) Do teachers ever experiment with their pupils?

6. (a) Have you ever been part of an experiment?
   (b) Are there any experiments going on in your school?

7. (a) Do you think that your arithmetic class is part of an experiment?
   (b) Why?

8. (a) Are you doing something that will help make the experiment a success?
   (b) What makes an experiment successful?

Analysis of the students' responses to this set of questions brought out several points. First, all questions in the interview had possible cue value, even those included originally to establish rapport. The questions appeared to vary in their cue value depending on their position in the interview schedule as well as differing in cue value through the question itself. Second, the addition of questions in order to secure clarification or to encourage elaboration, while increasing the flexibility of the interview technique, did increase the quantity of responses but they also tended to decrease the validity of responses because of an increase in demand characteristics by reinforcing certain responses. Third, the length of the student's attendance at the school influenced the responses to certain questions on the interview. Fourth, even though the subjects were known to have been continually cued from several sources about the team-teaching experiment, there were some interviews that elicited no indication of knowledge of participation in the project.

Scaling Awareness. The third pilot study was designed to serve two purposes. The first was to elicit responses directly related to awareness of experimental variables that were of a nature similar to those manipulated in the field research phase. The second was to compile a list of general statements from the responses made to the interviews in order to develop criteria measures for the evaluation of subject responses to the interviews of the field research phase of the project. It was anticipated that the development of the awareness measure would proceed by categorizing student responses into varying degrees or levels of awareness as determined by independent judgments. Numerical values would then be assigned to each of the categories ranging from zero for those students showing no awareness up to a highest value depending on the number of categories established by the judges. Such a procedure carried out for each of the experimental variables in the field research phase would result in criteria for an awareness measure for each of the changes in method, teacher, and cueing. Summing the three scores would give each subject a total awareness score.

The interview used was primarily the same as for the second pilot student with only minor revisions in questions necessary to accommodate the specific variables as noted in the question set below. Schools conducting educational experiments or innovation projects at the fifth grade level which employed, singularly or in combination, experimental variables similar to those manipulated in the present study were located. The "new method" or arithmetic variable was located at a school in a small urban community north of Columbus, Ohio. The different teacher variable was located at two elementary schools in

Dayton, Ohio. The cueing variable was located at three elementary schools in the Columbus system which were not participating in the field phase of the project.[2] The specific set of questions administered to student samples as before are given below.

Question Set III - Awareness of Experimentation at Elementary Schools Having Treatments of New Mathematics, Different Teacher. Purpose statement - what elementary pupils think about school.

1. Of all the grades - what was your favorite? Why?
2. In this grade what were your three most favorite subjects?
3. What is an experiment?
4. What kinds of people do experiments? Would you describe one?
5. Do teachers ever experiment?
6. Do schools ever experiment?
7. Are there any experiments going on in your school?
8. Are you part of this experiment?
9. How do you know you are part of this experiment?
10. What did you do in this experiment?
11. Because you were in the experiment, did you work harder?
    (a) How?
    (b) Why?
12. Why do you think they are doing this experiment?
13. (Improvise question about success and evaluation of the experiment within context of response to question 12.)

The specific set of questions for the students involved in the schools receiving direct cue about participation is given below.

Question Set IV - Awareness of Experimentation at Elementary Schools Where Students Were Receiving Direct Cues at Three and Six Week Intervals or Not at All. Purpose statement - to find out what students think about school.

---

[2] This sample of schools was drawn from a project run independently of the main project and was conducted by Dr. Lewis Coon. The students in the schools were given pre- and posttests in mathematics with one half of the schools being told at selected intervals that they were in an experiment and the other half not so told. The purpose was to determine if test administration plus some form of direct cueing would create some conditions of awareness. Summary data for this independent project are available from Dr. Coon.

1. Of all the grades in which you have been in school, what was your favorite one? Why?
2. Do you know what an experiment is?
   - (a) If yes - what is it?
   - (b) If no - ask about their familiarity with science experiments in school.
3. What kinds of people do experiments?
4. Do teachers ever experiment?
   - (a) If yes - How?
   - (b) If no - Proceed to #5.
5. Do schools ever experiment?
   - (a) If yes - How?
   - (b) If no - Proceed to #6.
6. Are there any experiments going on in this school?
   - (a) If yes - How do you know?
   - (b) If no - Are there any changes or differences in school this year as compared to previous years in school?
     1. If yes - What are they?
     2. If no - Terminate the interview.
7. Are you part of an experiment?
   - (a) If yes - How do you know?
   - (b) If no - Terminate the interview.
8. What did you do in this experiment?
9. Because you were in an experiment did you work any harder than you have before in school?
   - (a) If yes - How? and then Why?
   - (b) If no - Proceed to #10.
10. Why do you think this experiment is being done?
11. Do you think what you did in the experiment will have anything to do with the success of the experiment?
    - (a) If yes - Why?
    - (b) If no - Terminate the interview.

The student responses obtained from these three situations were categorized separately by project staff members[3] and a numerical scale developed. The resultant scale was used to score all the interviews accumulated during the pilot studies in order to evaluate the scale's effectiveness in discriminating levels of awareness.

---

[3] The responses to the new mathematics situation were analyzed by project staff members already familiar with awareness instrument development. The responses to the new mathematics situation were analyzed by project staff members who joined the project during this period but who were not familiar with the previous work on scale development and hence could be considered as naive judges. Responses to the cueing situation were analyzed by project staff members familiar with prior scale development procedures.

During the process of scoring the responses, several conditions emerged causing the project staff to consider that the oral interview approach was becoming impractical for this particular project for two principal reasons. First, the amount of time that would be required to interview and score responses for each of approximately two hundred subjects involved seemed prohibitive. Second, the scaled scoring procedure developed was not sensitive enough to indicate anything but a gross measure of being either aware or not aware of experimental participation. Although refinement might have been possible, the time necessary to continue with the scale development was not available relative to the progress of the field study and the total time left in the project.

Staff conferences considering these problems lead to a decision to try another approach. It was decided that the problems of available development and administration time and cumbersome scoring might be avoided by developing a paper-pencil instrument possessing an objective scoring scheme. The interview technique was thus discarded as not feasible for this project and preliminary plans for the development of a paper-pencil instrument were instituted.

Development of the Final Awareness Instrument

The basic format of the final scale developed to measure awareness employed the technique generally known as the method of summated ratings as developed by Likert (46) and further described by Edwards (24). This format was selected primarily because it showed promise of providing maximum information yet at the same time making minimum demands on the part of the subject. It was felt maximum information would be generated if the statements were precise and sensitive enough since the subjects' ages had to be taken into consideration. Minimum demands would be placed on the subjects since they would be required to circle only that response which best represents their reaction to the given statements. It should be emphasized, however, that in adopting this format there was no implication that awareness was an attitude. The definition of awareness used for this scale was similar to the one presented above.

Item Construction. Once the format decision was made, attention was given to determining the content of the individual statements. The words and sentences of each item had to be familiar and straightforward so as to render little or no difficulty in reading by the subjects at a third grade level. The words for the individual statements were, therefore, selected using the guidelines established by Thorndike and Lorge (78). Only simple in contrast to compound sentences were utilized. No assumptions were made regarding who had awareness and who did not. It was quite possible that "the slow reader" in the experimental sample had awareness. Therefore, the subject's reading skill should not prohibit him from expressing it.

The individual statements were focused on (1) the new arithmetic text, (2) the different teacher, and (3) classmates interaction. By so doing, recognition was given to the possibility that a subject may not have had awareness as a function of either a new arithmetic book or a different teacher but rather as a function of fellow classmates. A fourth focus was awareness in general which meant here awareness as a function of the interactions of the new arithmetic book, the different teacher, and classmate interactions. Examples of individual statements so constructed are presented below.

A. From the arithmetic material we use, it is clear that we are part of an experiment.

B. There is something special about our arithmetic teacher.

C. My classmates think our class is taking part in an experiment.

D. Our teacher is teaching us our school work and is also trying to find out something special about us.

Consideration was given to the fact that the individual statements must be sensitive in order to measure the variable of interest--but at the same time, should not themselves generate awareness. Several neutral items, therefore, were constructed in an attempt to control for this possible source of error. The neutral items focused on other subjects such as reading, writing, spelling, and science. Several items dealing with social adjustment were subsequently included for this reason as well as for others as noted below. Several negative statements were utilized to reduce the likelihood of a subject establishing a response set. The use of such negative statements was limited due to the content of the response categories.

The design of the response categories was predicated by the dependent variable. It was desired to know to what extent a given subject had awareness. The initial response categories developed are shown below with an illustrative example.

"I am part of an experiment during this school year."

I know this is true	I suspect this is true	I cannot decide	I suspect this is not true	I know this is not true

While the use of the negation "not true" limited the use of negative statements in order to avoid double negatives it was felt that such phrasing would be more easily understood than the word "false" and was therefore employed. According to Thorndike and Lorge the word "suspect" was determined to be sufficiently familiar enough to fourth-grade subjects to justify its use. The response categories were assigned integral rates with "high" awareness given a rating of 5 and "low" a rating of 1.

A total of 28 items relating to awareness plus a total of 14 neutral items was administered to a sample of approximately 60 fifth-grade subjects for the primary purpose of determining whether or not the words and sentences created any difficulty. Results of this study showed that words such as "suspect" or "experiment" were not too difficult for subjects to understand. Some subjects tended to contradict themselves with double negatives, however, thereby indicating some difficulty with them.

In view of the findings on this preliminary trial, the following response categories were designed.

I know this is true	I suspect this is true	I cannot decide	I suspect this is false	I know this is false

These response categories were considered to be the most straightforward and suitable for purposes of the study.

Although the above response categories were considered satisfactory by the project staff, some objection to the use of the word "suspect" was voiced by the school system from which the sample was taken. Although the project staff met and discussed the problem with local school officials, no precise reason for the objection was determinable. The project staff was told, however, that unless an appropriate modification was made it would not be possible to administer the scale to the subjects involved in the experiment. It was decided after careful consideration of possible alternatives that a suitable equivalent of "I suspect..." was "I think...or I am not sure." The following response categories were therefore employed for the final instrument.

I know this is true	I think this is true but I am not sure	I cannot decide	I think this is false but I am not sure	I know this is false

Since there was a great deal of interest in determining whether or not subjects knew they were in an experiment, it was deemed worthwhile to know what fifth-graders understood by the word "experiments." Such information would be helpful in designing statements which included phrases that are used by the subjects themselves. Such information would also aid in determining the various meanings the word has for the subjects and hence assist in eliminating those meanings which were of no interest to the project. In an interview situation, subjects were asked "What is an experiment?" Some examples of the responses are given below.

A. Seeing if something will work (What do you mean?) Well, in your kind of experiment you are going to see which method of arithmetic works best, isn't it?

B. (Do you think teachers can do experiments?) Yes, on kids and stuff in the classroom.

C. (How would teachers do experiments?) They try to find new ways to teach the kids...or (Why do you think this experiment is being done?) To find out how to teach the children arithmetic easier and it wouldn't be as difficult.

D. We had a lot of experiments out of a science book.

E. When you try to find out something like how much water is in an orange or something.

F. Like when you are in a laboratory or something you have all kinds of chemicals drop in to see if it will blow up.

Of the above examples, items A, B and C were considered appropriate responses in terms of the study while examples D, E and F point out the kind of experiment definition in which the project staff was not interested. Statements that allow for these other interpretations were avoided. From the three examples (A, B and

C, above) the following statements were designed for inclusion in the scale:

A. A school is trying to see which method of arithmetic works best.

B. The school is trying to find new ways to teach children arithmetic easily.

C. Our school is doing an experiment on us kids in the classroom.

It was necessary that the items be free from ambiguity. In order to accomplish this goal, it was necessary that the project staff think and re-think as fifth-graders. What may appear to be a perfectly clear statement to the staff often is interpreted in various ways by subjects at this level. Each item was examined and re-examined and new ones designed as necessary. Some items were eliminated. For example,

A. In responding to the statement, "I am part of an experiment during the school year." it would be expected that a subject who had awareness would respond affirmatively to this statement. However, if the subject were to understand "experiment" as in the illustrations D, E and F above, a subject who did not have awareness would also affirm it. The statement was therefore re-structured to read as follows: "I am part of an experiment in arithmetic this year."

The above item was selected for illustration because it was felt to have special significance. The project staff considered it to be a key statement although such a decision was made on a priority grounds. The statement contains the essential of what we wanted to measure. The total effort in scale development centers on asking the same question in 43 different ways. An example of an item eliminated is as follows:

A. "There is something special about our arithmetic teacher." If the subject's awareness was a function of his different arithmetic teacher, the response should be affirmative. A subject who did not have awareness could also respond positively to this statement. For example, his teacher could be different because she dressed in an unusual manner. Such items were eliminated.

In addition to the reasons cited earlier, several items dealing with social adjustment were included in order to avoid establishment of a response set (i.e., they demanded a negative response) and to determine whether or not subject's performance on this scale was a functional adjustment to the school's situation. The items used for this part of the scale were drawn from a standardized personality instrument[4] designed for measuring social adjustment at the elementary school level.

A preliminary form of the scale (identified by the project staff as Form 2 AH) was developed. It consisted of 43 items relating to awareness, 11 neutral items, and 24 items dealing with social adjustment. This preliminary scale was administered to a total of 52 subjects drawn at random from four of the eight treatment combinations[5] in the experiment at a period of time approximately three months before the end of the experiment or the administration of the final posttest. The project staff discussed at some length the desirability of utilizing a small group of subjects from each of the several treatment conditions in order to develop a scale which in turn would be used as subsequent measuring of awareness within these same subjects. That is, the very act of administering the preliminary tryout scale to the subjects in the experiment might well create the very effect in which there was interest. In view of the absence of comparable subjects on which to try out this scale, a decision to accept this risk on the assumption that the time intervening between the tryout and final scale might diminish any curiously aroused. It was also decided that it would be possible to determine, through a separate analysis, whether or not the average awareness scores for those subjects used in the preliminary tryout would be significantly different from those constituting the balance of the sample.

Upon completion of this preliminary administration, a principal-factor analysis was performed on the data for the 43 awareness items followed by a Kaiser Varimox rotation. Of the total variance, 66 percent was accounted for by this analysis.

---

[4] Selected items were drawn from The California Test of Personality, 1953 Revision, with special permission of the publishers.

[5] The specific treatments utilized were (1) Regular-Conv-Not Told, (2) Regular-SMSG-Not Told, (3) Different-Conv-Not Told, and (4) Different-SMSG-Not Told.

The items with factor loadings of no less than 38 around each factor were then analyzed. A total of ten factors were identified. No attempt was made to apply descriptive names to them. The factor loadings for the 43 awareness items across the ten factors are presented in Table B-1. It can be observed that 23 awareness items appear in Factor 1. If consideration is given only to those items with heavy loadings in only one factor, a matrix such as that shown in B-2 is derived. In this table it can be observed that 12 of the 28 items appear in Factor 1. It is also evident that the remaining items are well dispersed across the other nine factors. It appeared therefore that the data reflected a single best factor, namely Factor 1. Each of the ten factors along with the best characterization of a given factor is presented below.

Factor 1. This factor reflects personal involvement in the action described by the statements. The whole class is involved in this action as well. Further, it is evident that the action focuses on the arithmetic material and the arithmetic teacher. The key item described earlier appears in this factor and in no other and has the heaviest factor loading.

Factor 2. A knowledge of experimentation by the schools is indicated with some orientation to arithmetic and the materials being used. However, personal involvement in the action described is not evident.

Factor 3. The school is participating in an experiment, however, only general action is described in this factor.

Factor 4. The dominant characteristic here is the subject's mindfulness of a special activity taking place, with some focus on arithmetic. This special activity is for the most part quite general.

Factor 5. This school is doing an experiment and the subject sees himself as a part of it but no specific focus however is indicated.

Factor 6. The subject sees his class as part of an experiment but this does not make it different from ordinary classes.

Table B-1.--Matrix of Factor Loadings of Items Across Ten Factors

Item	Factor										Total
	1	2	3	4	5	6	7	8	9	10	
1	-	.7423	-	-	-	-	-	-	-	-	1
2	.7097	-	-	-	-	-	-	.2912	.2797	-	3
3	.3373	-	-.6950	-	-	-	-	-	-	-	2
4	.6156	-	-	-	.2978	-	-.3045	-	-	-	3
5	-	-	-	-	-	-	-	-	.6599	-	1
6	.3988	-	-	-	-	-	-	-	-	-	1
7	-	-	-.7605	-	-	-	-	-	-	-	1
8	-	-	-	.6555	-	-	-	-	-	-	1
9	.4344	-	-	-	-	-	-	.6505	-	-	2
10	.6239	.3373	-	-	-	-	-	-	-	-	2
11	-	-	-.3916	-	.6124	-	-	-	-	-	2
12	.6025	-	-	-	-	-	-	.3219	-	-	2
13	-	-	-	-	-	-.6296	-	-	-	-	1
14	-	-	.4016	-	-	.2864	-	-	-	-.4796	3
15	-	.5947	-	-	-	-	-	.3515	-	-	2
16	-	.6410	-	-	-	-	-	-	.3062	-	2
17	.2846	-	-	-	-	-	-	-	-	.6491	2
18	.5809	-	-.4735	-	-	-	-	-	-	-	2
19	-	.4041	-	-	-	-	-	.5670	.4280	-	3
20	-	.5569	-	-	.3798	-	-	.3217	-	-	3
21	-	-	-	-	.6738	-	-	-	-	-	1
22	-	-	-	.6836	-	-	-	-	.2969	-	2
23	-	-	-	.3238	-	-	.2913	-	-	.3926	3
24	.4820	-	-	.2848	-	-	-	-	-	.2973	3
25	-	-	-	-	-	-	-.3032	-	-	-	1
26	-	-	-	-	-	-	-.8002	-	-	-	1

Continued on next page

Table B-1.--Matrix of Factor Loadings of Items Across Ten Factors (Continued)

Item	Factor										Total
	1	2	3	4	5	6	7	8	9	10	
27	.3307	-	-	.2994	.5657	-	.4425	-	-	-	4
28	.6592	-	-	.4144	-	-	-	-	-	-	2
29	.3136	.4368	-	-	-	-	-	-	.3822	-	3
30	.5229	-	-.3116	.3361	-	-	-	-	-	.3225	4
31	.3026	.4710	-	-	-	-	-	.5572	-	-	3
32	.3646	-	-	.3429	-	-.2773	-	-	.3160	.3875	5
33	.7575	-	-	-	-	-	-	-	-	-	1
34	.4491	-	-.5329	-	-	-	-	-	-	-	2
35	.7275	-	-	.3335	-	-	-	-	-	.3040	3
36	-	-	-	.5273	-	-	-	-	-	.4238	2
37	-	-	-	-	-.5240	-	-	-	-	-	1
38	.6596	-	-	-	-	-	-	-	-	.3512	2
39	.4545	-	-	.4994	-	-.4532	-	-	-	-	3
40	-.3524	-	-	-	-	.7170	-	-	-	-	2
41	.5830	-	-.3107	-	-	-	-	-	-	.3584	3
42	-	-	.5049	-	-	.5138	-	-	-	-	2
43	-	-	-	-	-	-.3487	-	-	-	.2763	2
Total	23	8	9	11	6	7	5	7	7	11	94

Table B-2.--Matrix of Factor Loadings of Items in Only One Factor for Ten Factors

Item	Factor										Total
	1	2	3	4	5	6	7	8	9	10	
1	-	.7423	-	-	-	-	-	-	-	-	1
2	.7079	-	-	-	-	-	-	-	-	-	1
3	-	-	-.6950	-	-	-	-	-	-	-	1
4	.6156	-	-	-	-	-	-	-	-	-	1
5	-	-	-	-	-	-	-	-	.6599	-	1
6	.3988	-	-	-	-	-	-	-	-	-	1
7	-	-	-.7605	-	-	-	-	-	-	-	1
8	-	-	-	.6555	-	-	-	-	-	-	1
9	-	-	-	-	-	-	-	-	-	-	0
10	.6239	-	-	-	-	-	-	-	-	-	1
11	-	-	-	-	.6124	-	-	-	-	-	1
12	.6025	-	-	-	-	-	-	-	-	-	1
13	-	-	-	-	-	-.6296	-	-	-	-	1
14	-	-	-	-	-	-	-	-	-	-	0
15	-	.5947	-	-	-	-	-	-	-	-	1
16	-	.6410	-	-	-	-	-	-	-	-	1
17	-	-	-	-	-	-	-	-	-	.6491	1
18	-	-	-	-	-	-	-	-	-	-	0
19	-	-	-	-	-	-	-	-	-	-	0
20	-	.5569	-	-	-	-	-	-	-	-	1
21	-	-	-	-	.6738	-	-	-	-	-	1
22	-	-	-	.6836	-	-	-	-	-	-	1
23	-	-	-	-	-	-	-	-	-	-	0
24	.4820	-	-	-	-	-	-	-	-	-	1
25	-	-	-	-	-	-	-	-	-	-	0
26	-	-	-	-	-	-	-.8002	-	-	-	1

Continued on next page

Table B-2.--Matrix of Factor Loadings of Items in Only One Factor for Ten Factors (Continued)

Item	Factor										Total
	1	2	3	4	5	6	7	8	9	10	
27	-	-	-	-	-	-	-	-	-	-	0
28	.6592	-	-	-	-	-	-	-	-	-	1
29	-	-	-	-	-	-	-	-	-	-	0
30	.5229	-	-	-	-	-	-	-	-	-	1
31	-	-	-	-	-	-	-	-	-	-	0
32	-	-	-	-	-	-	-	-	-	-	0
33	.7575	-	-	-	-	-	-	-	-	-	1
34	-	-	-	-	-	-	-	-	-	-	0
35	.7275	-	-	-	-	-	-	-	-	-	1
36	-	-	-	-	-	-	-	-	-	-	0
37	-	-	-	-	-.5240	-	-	-	-	-	1
38	.6596	-	-	-	-	-	-	-	-	-	1
39	-	-	-	-	-	-	-	-	-	-	0
40	-	-	-	-	-	.7170	-	-	-	-	1
41	.5830	-	-	-	-	-	-	-	-	-	1
42	-	-	-	-	-	-	-	-	-	-	0
43	-	-	-	-	-	-	-	-	-	-	0
Total	12	4	2	2	3	2	1	0	1	1	28

Factor 7. Insufficient information did not permit identification of this factor.

Factor 8. This school has a special interest in arithmetic and the subject sees a different teacher as having a vital role in this interest.

Factor 9. This school has a special interest in arithmetic but the subject may perceive this as being independent of any experimentation.

Factor 10. This school is doing an experiment in arithmetic and the subject sees his teacher as having a part in it.

Of the above ten factors, Factor 1 is the most specific bringing its essence into sharp focus. It focused on the variable interest, awareness, whereas the other factors were more general and less definitive.

Using the criteria of (a) factor identification, (b) the extent to which a given item appeared across factors, and (c) the factor loading of items in a given factor, a total of 17 items were selected for inclusion in the final awareness instrument as described below.

From Table B-1 it can also be seen that while items 18, 39, 32, 34, and 9 have loadings in more than one factor they nevertheless have substantial loadings in the factor of interest, and therefore, could be put to good use. Combining these five items with the 12 items from Table B-2, a total of 17 items was available to measure the variable of interest. Table B-3 presents the final 17 items along with their factor loadings.

The final awareness scale, (designated as Form 3 AH) consisting of the 17 awareness items plus 9 neutral items and 24 items of social adjustment arranged in random order was administered to all subjects (N = 342) in the several treatment groups within a one-week period after the final posttest was administered. Because there was even the possibility that the treatment group receiving no modifications in terms of teacher, textbook or being informed might possibly have developed some awareness, it was decided to administer the scale to an external control at the same level which had not been involved in the experimental situation. Consequently, three classes were randomly selected at the fifth-grade level with a total N of 55 subjects in order to make such comparisons.

Table B-3.--The Final Seventeen Awareness Items with Their Factor Loadings

Factor Loading	Awareness Item
.7575	I am part of an experiment in arithmetic this year.
.7275	Our arithmetic teacher is taking part in an experiment the school is doing.
.7097	Our whole class is part of an experiment in arithmetic this year.
.6596	From the arithmetic material we use it is clear that we are part of an experiment.
.6592	Our arithmetic material is part of an experiment the school is doing.
.6239	Our arithmetic material is different than ordinary arithmetic material for a very special reason.
.6156	From the way we study arithmetic it is clear that the school is doing an experiment on us kids.
.6025	We have a different arithmetic teacher for a very special purpose.
.5830	When we study arithmetic it is clear that the school is doing an experiment on us kids.
.5809	Our teacher is taking part in an experiment in arithmetic this year.
.5229	My classmates think our teacher is doing an experiment on us kids.
.4820	My classmates think our whole class is part of an experiment the school is doing.
.4545	Our class is not an ordinary fifth grade class since we are part of an experiment the school is doing.
.4491	Our teacher is trying to find out something about fifth graders for a very special reason.
.4344	From the different arithmetic teacher we have it is clear that we are part of an experiment.
.3988	The school is trying to find out something special about arithmetic.
.3646	When the teacher teaches arithmetic it is clear that she is doing an experiment on us kids.

Results and Discussion

Two types of scores were used in analyzing the data. The first score was a summation of the ratings assigned by the subject to the individual 17 statements. The second score was the mean scale score derived by dividing the summated score by the total number of items. Since individual subject scores were not of interest in the study at this point, the results are presented in terms of treatment groups. Table B-4 presents the mean summated in scale scores according to treatment groups.

Table B-4.--Mean Summated and Scaled Awareness Scores

TREATMENT	N	MEAN SUMMATED SCORES	VARIANCE	MEAN SCALE SCORE
Regular-Conv-Told	32	53.19	221.31	3.13
Regular-Conv-Not Told	49	34.04	132.45	2.00
Regular-SMSG-Told	40	69.50	96.36	4.09
Regular-SMSG-Not Told	54	54.52	174.85	3.21
Different-Conv-Told	41	67.98	152.17	4.00
Different-Conv-Not Told	36	48.03	180.26	2.83
Different-SMSG-Told	48	70.15	77.62	4.13
Different-SMSG-Not Told	42	72.74	55.17	4.28

It can be observed in this table that the treatment group having a Different teacher and the SMSG materials but not being told had both highest mean summated score and mean scale score. The group having the Regular teacher with the Conventional materials and Not Told received both the lowest mean summated score and the lowest mean scale score. A comparison of the mean scale scores for the experimental internal control group (Regular-Conv-Not Told) and the external control group mean scale scores revealed that these were very similar.

The performance of the internal control group did not differ significantly (probability less than .05) from the external control group on the awareness scale. This finding gives further reliability to the performance of the internal control group.

A simple analysis of variance utilizing the mean summated scores as criteria measures was made. The results appear in Table B-5. The F-value was 62.73 which is significant at greater than the .01 level (7,334 d.f.). To determine which differences between treatment means were significant, the Newman-Keuls technique of testing between individual treatment means as outlined by Winer (81) was employed. The results of this analysis are presented in Table B-6. A total of 19 significant differences at greater than the .01 level were obtained. Each of the seven treatment groups had higher awareness scores significantly different from the control group. From these results it would appear evident that each of the variables (Teacher, Method, Cueing) were in themselves sufficient to generate awareness. Inspection of the results reveals that there is a slight trend existing which indicates that being informed directly or working with a special arithmetic text may be stronger sources of awareness than having a Different arithmetic teacher. This finding however would have to be substantiated by further research. Results also indicate that a combination of the variables generated more awareness than did any single variable. Similar results were obtained by Lawrence (44) with regard to the last finding using a less refined awareness measure coupled with a shorter experimentation period.

Table B-5.--Summary Table for Analysis of Variance

Source	Sum of Squares	d.f.	Mean Squares	F
Treatment	58,565.35	7	8,366.48	62.73**
Error	44,550.32	334	133.38	
Total	103,115.67	341		

An examination of the awareness items to which both the external and internal control groups responded with low awareness revealed that such items likewise elicited low awareness responses from the seven treatment groups. Conversely, items to which the treatment groups responded with high awareness also elicited high awareness responses from the control groups.

Table B-6.--Tests on Differences Between All Pairs of Means

Treatments		1	2	3	4	5	6	7	8
	Means	34.04	48.03	53.19	54.52	67.98	69.50	70.15	72.74
Regular-Conv-Told	34.04	---	13.99**	19.15**	20.48**	33.94**	35.46**	36.11**	38.70**
Regular-Conv-Not Told	48.03		---	5.16	6.49	19.15**	21.47**	22.12**	24.71**
Regular-SMSG-Told	53.19			---	1.33	14.79**	16.31**	16.96**	19.55**
Regular-SMSG-Not Told	54.52				---	13.46**	14.98**	15.63**	18.22**
Different-Conv-Told	67.98					---	1.52	2.17	4.76
Different-Conv-Not Told	69.50						---	.65	3.24
Different-SMSG-Told	70.15							---	2.59
Different-SMSG-Not Told	72.74								---

**Difference significant at 1 percent level

An example of the latter is the statement "Our school is trying to see which method of arithmetic works best." To a subject in either control group, this statement may very well be true. Nevertheless, it has nothing to do with an experiment in arithmetic.

As noted above, several items relating to social adjustment were included as part of the total awareness inventory to determine if a subject's performance on the awareness scale was a function of her adjustment to school. Pearson product-moment coefficients of correlation between awareness item scores and social adjustment scores were of very low magnitude indicating that a given subject's performance on the awareness scale was not a function of his adjustment to the school situation.

In addition to the social adjustment analysis, data were secured from each subject with regard to mental ability as measured by the California Test of Mental Maturity available in project files. A correlational analysis indicated that a subject's awareness was not a function of his intelligence. From this finding, it can be observed that a subject with low intelligence can have awareness. The influence that awareness has on a bright student in contrast to the less able student would constitute another problem for research.

In summary, it should be emphasized that the nature of this study was innovative. Should researchers desire to replicate the work, it is suggested that additional new items be constructed in the design of a preliminary scale and that this scale be administered to at least 200 subjects in order to reduce error before a factor analysis is performed on the data.

I

FORM 3AH

DIRECTIONS:

Please respond to ALL statements. Put a circle around ONE AND ONLY ONE response to each statement.

There is NO RIGHT OR WRONG response. We want your honest reaction to each and every statement. Circle the response which best describes your feelings about each statement.

NAME____AGE__________

DATE OF BIRTH______________________MALE_______FEMALE_______

GRADE_________SCHOOL_______________________________________

5/13/65

1. I am part of an experiment in reading this year.[6]

I know this is true	I think this is true but I am not sure	I cannot decide	I think this is false but I am not sure	I know this is false

2. People are often so unfair that I lose my temper.

3. There is a very special purpose for the way we practice writing.

4. School work is so hard that I am afraid I will fail.

5. Sometimes I need something so much that it is all right to take it.

6. Our arithmetic teacher is taking part in an experiment the school is doing.

7. Our class is an ordinary fifth grade class.

8. From the way we study arithmetic it is clear that the school is doing an experiment on us kids.

9. I often make my classmates do things they don't want to.

10. The boys and girls at school like me as well as they should.

11. The school is taking part in an experiment in arithmetic this year.

12. Our spelling material is different than ordinary spelling material for a very special reason.

13. The writing material we use has a very special purpose.

14. We have a different arithmetic teacher for a very special purpose.

15. Our school is trying to see which method of arithmetic works best.

16. People often act so mean that I have to be nasty to them.

---

[6] Since the response terms were the same for all items, they have been omitted for the sake of brevity.

17. When the teacher teaches arithmetic it is clear that she is doing an experiment on us kids.

18. Our school is trying to see if there is a new and better way of doing arithmetic.

19. I like to scare or push smaller boys and girls.

20. Some of the teachers "have it in for" pupils.

21. It is hard to make people remember how well I can do things.

22. The school is trying to find out something special about spelling.

23. The boys and girls at school are usually nice to me.

24. When we study arithmetic it is clear that the school is doing an experiment on us kids.

25. Unfair people have often said that I made trouble for them.

26. There is something fishy about the way we practice writing.

27. Someone at home is so mean that I often have to quarrel.

28. I am part of an experiment in arithmetic this year.

29. Our spelling material is part of an experiment the school is doing.

30. Our school is not doing an experiment on us kids in the classroom.

31. People often ask me to do such hard or foolish things that I won't do them.

32. I often have to make a "fuss" or "act up" to get what I deserve.

33. My schoolmates seem to think that I am nice to them.

34. Some of the teachers do not like to be with the boys and girls.

35. I would like to stay home from school a lot if it were right to do so.

36. From the different arithmetic teacher we have it is clear that we are part of an experiment.

37. Many of the other boys or girls claim that they play games more fairly than I do.

38. My classmates think our whole class is part of an experiment the school is doing.

39 Many of the children get along with the teacher much better than I do.

40. The school is trying to find out something special about arithmetic.

41. Someone at school is so mean that I tear, or cut, or break things.

42. Our arithmetic material is part of an experiment the school is doing.

43. Our teacher is trying to find out something about fifth graders for a very special reason.

44. Our class is especially different than ordinary fifth grade classes since we are part of an experiment the school is doing.

45. The school is trying to find new ways to teach children arithmetic easily.

46. From the arithmetic material we use it is clear that we are part of an experiment.

47. Our whole class is part of an experiment in arithmetic this year.

48. Most of the boys and girls at school are so bad that I try to stay away from them.

49. Our teacher is taking part in an experiment in arithmetic this year.

50. My classmates think our teacher is doing an experiment on us kids.

51. It is fun to do nice things for some of the other boys or girls.

52. Classmates often quarrel with me.

53. Our arithmetic material is different than ordinary arithmetic material for a very special reason.

54. The children would be happier if the teacher were not so strict.

55. Our class is not an ordinary fifth grade class since we are part of an experiment the school is doing.

56. The school is taking part in an experiment in reading this year.

Appendix C--Part I

Major Areas of Research (Divided According to Four Special Categories), Subgroups within Major Areas (When Applicable) and Education Levels Applicable to Major Areas[1]

Subject Matter Areas

Adult Education (Level 4 only) - subgroups involve institutions offering programs.

- Churches
- Colleges
- Cooperative Agricultural Extension Services
- Correspondence Schools
- Evening Colleges
- General Universities
- Group Work Agencies (YMCA, etc.)
- Labor Unions
- Libraries
- Museums
- Public (Parochial) School Systems

Communication Arts (All Levels) - no subgroups

English (Levels 3 and 4 only) - subgroups involve fundamental elements of study.

- Composition
- Grammar
- Language
- Literature

Foreign Language (All Levels) - subgroups involve specific foreign languages.

- French
- German
- Greek
- Latin
- Russian
- Spanish

Handwriting (All Levels) - subgroups involve fundamental types.

---

[1] See last page in this table for special notation used to indicate levels of education used for major areas.

Cursive
Manuscript

Health Education (All Levels)

No subgroups

Home Economics (All Levels)

No subgroups

Industrial Arts (All Levels) - subgroups involve shop activities.

Crafts (leather, ceramics, etc.)
Drafting
Electricity (Electronics)
Graphic Arts
Metal Working
Power Mechanics
Wood Working

In-Service Education (Level 4 only) - subgroups involve sponsoring institutions.

Colleges and Universities
Government
Public (Parochial) Schools

Mathematics (All Levels) - subgroups involve specific courses.

Algebra (I and II)
Arithmetic (elementary level only)
Business Mathematics
Calculus
General Mathematics
Geometry (Plane, Solid, Analytical)
Terminal Mathematics
Trigonometry

Music (All Levels)

No subgroups

Physical Education (All Levels)

No subgroups

Professional Education (Level 4 only) - subgroups involve specific professions.

- Architecture
- Business
- Dentistry
- Engineering
- Home Economics
- Law
- Library Science
- Medicine
- Nursing
- Pharmacy
- Social Work
- Teaching
- Theology

Reading (All Levels) - subgroups involve specific areas within the field.

- Organization of Reading Instruction
- Reading Improvement Programs
- Stimulating Interests and Improving Tastes
- Reading in the Content Fields
- Reading Materials
- Reading Readiness

Science (All Levels) - subgroups involve specific courses.

- Anatomy
- Astronomy
- Biology
- Botany
- Chemistry
- Elementary Science
- General Science
- Geology
- Physics
- Physiology
- Zoology

Social Sciences (Studies) - (All Levels) - subgroups involve specific courses.

- American History
- Civics
- Current Events
- Economics

Geography
Government
Problems Courses
Psychology
Sociology
World History (Culture)

Spelling (All Levels)

No subgroups

Vocational Education (All Levels) - subgroups involve relationship of education and work.

- Education prior to work experience
- Vocational Agriculture
- Business Education
- Vocational Training for Industry
- On-The-Job Training

Facilitative Devices or Procedures for Improving Teaching

Audio-Visual Communication (All Levels) - subgroups involve special devices and methods.

- Field Trips
- Filmstrips and Slides
- Graphic Materials
- Motion Pictures
- Pictorial Illustrations
- Radio
- Recordings
- Television
- Three Dimensional Materials

Curriculum Organization (All Levels) - subgroups involve both general school organization and special school programs.

- General Types of School Organization
- Grouping
- Provisions for Exceptional Children

Training Devices (All Levels) - subgroups involve specific devices.

- Programmed Texts
- Teaching Machines
- Training Machines (Flight trainers, etc.)

General Teaching and Learning Methods

Higher Mental Processes (All Levels) - subgroups involve basic cognative processes.

Associative
Creative
Critical
Problem-solving

Methods of Teaching (All Levels) - subgroups involve classroom activities and specific studies employing multiple subject-matters.

Individualizing Instruction
Inter-personal Relations
Patterns of Instruction
Specific Activities in Teaching
Multiple-Subject Matter

Readiness (All Levels) - subgroups involve primary factors in readiness for any school activity.

Emotional Adjustment
Experience
Maturation

Areas of Special Interest For the Hawthorne Effect Project

Experimental Design (No Levels)

No subgroups

Hawthorne Effect (No Levels)

No subgroups

Special Notation Used to Indicate Education Levels for Individual Major Areas

1. Pre-Primary (Kindergarten, etc.)
2. Primary
3. Secondary
4. Post-Secondary (College, Armed Forces, Community programs, in-service education, etc.)

Appendix C--Part 2

Abstract Summary Form
March 19, 1965

1. Abstract No. ____________
2. Major Area ____________________
3. Sub-group ____________________
4. Education Level ____________________
5. Experimental Variable (s) ____________________________
6. Comparison Condition (s) ____________________________
7. Group Differences on Criterion Measure (s):
   a. Experimental > Comparison ______
   b. Experimental = Comparison ______
   c. Experimental < Comparison ______
8. Explicit controls established for the "Hawthorne effect" or its possible components. yes no
9. Mention made of special interest or negative reaction of pupils. reported not reported
10. Bias or reaction on the part of teachers (or experimenter) mentioned. reported not reported
11. Experimental (new) method identified as one sometimes used as a comparison (old) method. yes no
12. Counter-balancing or group rotation design. yes no
13. Duration: ______________________________________
14. Degree of Mechanization: Experimental _________________
    Comparison _________________
15. Similarity in design to Hawthorne Effect Project's field experiment design:
   a. New vs. on-going teaching methods __________
   b. Regular vs. different teachers __________
   c. Explicit cueing concerning experimental participation __________
16. Experimental and control in same school. not reported yes no
17. Same teacher teaching experimental and comparison methods. not reported yes no
18. Groups equated by _______________
19. N for study ____________________

Appendix C--Part 3

Abstracting Procedure - Experimental Teaching Methods Studies

Step 1 - Categorizing the Study

1. The classification of a study to a major area of research and a relevant subgroup within that major area is based upon the experimental variable ("new" teaching methods) manipulated in the study.

2. The classification of the study to a level of education is based upon the grade level or age of the experimental population.

Step 2 - Reference Title

1. Transcribe all information concerning the investigator, title of research, etc., according to the specific bibliographic form applicable to the form of reporting.

Step 3 - Experimental Variable (s)

State the "new" teaching method (s) manipulated by the investigator.

Step 4 - Comparison Conditions

State the "on-going" teaching method (s) used with the comparison group (s) in the study.

Step 5 - Dependent Variable (s)

1. State that task, process or activity evaluated by the investigator which is intended to show the effects of experimental variable manipulation.

2. The dependent variable is not equivalent to the instrument used to measure the dependent variable, but may be used by the investigator to operationally define the dependent variable.

Step 6 - Experimental Population

1. Describe the general population from which the experimental sample was drawn using such factors as the names, sizes, locales, and types of institution

or school involved, and the age, sex composition and grade level of the general population.

2. Describe the experimental sample of subjects drawn from this general population for use in the study using such factors as the number of experimental and control groups, number and sex composition of subjects per group, and total number (N) of subjects in the experiment.

Step 7 - Equating Groups

1. Determine what method the investigator used in equating the experimental and control groups:

   a. matching group subjects on co-variables

   b. random assignment of subjects to groups used

   c. combination of both of the above

2. If a matching technique was used, specify the covariables on which subjects were matched, i.e., age, sex, IQ, subject matter achievement, socio-economic status, etc.

Step 8 - Duration

State the inclusive dates of the experiment and the total time involved in the study.

Step 9 - Instrumentation

1. State the tests or instruments used to measure each dependent variable.

2. Describe measures as to being standardized or specifically developed for the study in question.

3. Specify pre- and posttests.

Step 10 - Group Differences on the Criterion Measure of the Dependent Variable

1. State the differences between groups that were compared in terms of achievement scores, gain scores, etc., according to the specific instruments used as criterion measures.

Step 11 - Investigator's Conclusions

1. Quote from the study the investigator's statements or conclusions relevant to the hypotheses tested and the results achieved.

2. Disregard statements not relevant to the direct results of the experiment.

3. Quote any statements by the investigator concerning the Hawthorne effect or discussion of uncontrolled variables influencing or affecting experimental results. If such statements are made, write another locator card and another abstract according to the procedure described for Hawthorne effect of Experimental Design references outlined below.

Step 12 - Abstractor's Evaluation

1. State judgment regarding compatability of investigator's conclusions and reported data.

2. State observations made by the abstractor (but not by the investigator) of the possible influence on experimental results of such uncontrolled variables as the tentatively defined components of the Hawthorne effect, (attention, novelty, role expectation, and knowledge of results).

Step 13 - Abstractor's Identification

Write signature and the date the abstract was completed.

Abstracting Procedure - Hawthorne Effect or Experimental Design References

Step 1 - Major Area

State the major area, either Hawthorne effect or Experimental Design.

Step 2 - Reference Title

Describe the reference following the same procedure outlined for Step 2 of the Experimental Teaching Methods Abstract Format.

Step 3 - Relevant Discussion

1. The purpose here is to judge what discussion in the article or report is relevant for the project's use (see descriptions of the major areas in Table II to evaluate relevancy of discussion).

2. Copy verbatim that discussion deemed relevant employing standard quotation procedures.

Step 4 - Abstractor's Evaluation

State the reasons why this material appears to be relevant and its relation to the purposes of the Hawthorne effect Project.

Step 5 - Abstractor's Identification

In examining materials relevant to the Hawthorne effect or Experimental Design major areas it may be apparent that *most or all* of the discussion in a particular report or article is of value. *In this event, do not abstract this material*. Instead, do one of the following whichever is more convenient:

1. Make arrangements with the Project Director to have the material photocopied.

2. Write to the appropriate source for a reprint. (Follow procedure regarding all requested materials as described in Section VIII).